UNIVERSITY *of* TASMANIA
AUSTRALIA

Object-based remotely sensed image classification for identification of lowland native grassland communities in the Tasmanian Midlands

by

Bethany Melville

HBSc (Geography), Cert. MapSci

Submitted in fulfilment of the requirements for the Degree of Doctorate of Philosophy
School of Land and Food
University of Tasmania

This thesis contains no material which has been accepted for a degree or diploma by the University or any other institution, except by way of background information and duly acknowledged in the thesis, and to the best of my knowledge and belief no material previously published or written by another person except where due acknowledgement is made in the text of the thesis, nor does the thesis contain any material that infringes copyright.

Bethany Melville

14 November 2016

Authority of Access

Statement of Co-Authorship

The following people and institutions contributed to the publication of work undertaken as part of this thesis:

Bethany Melville, School of Land and Food, University of Tasmania
Arko Lucieer, School of Land and Food, University of Tasmania
Jagannath Aryal, School of Land and Food, University of Tasmania

Author details and their roles:

Chapter 2: Melville (90%), Lucieer (5%), Aryal (5%)

All authors contributed to design and development. Melville performed all analysis and writing. Lucieer and Aryal provided editorial advice.

Chapter 3: Melville (90%), Lucieer (5%), Aryal (5%)

All authors contributed to design and development. Melville performed all analysis, field work and writing of chapter. Lucieer and Aryal provided editorial advice.

Chapter 4: Melville (90%), Lucieer (5%), Aryal (5%)
All authors contributed to design and development. Melville performed all analysis and writing of chapter. Lucieer provided help with implementation of Python programming, and Aryal provided advice on statistical techniques. Lucieer and Aryal additionally provided editorial advice.

Chapter 4: Melville (85%), Lucieer (10%), Aryal (5%)
All authors contributed to design and development. Melville and Lucieer performed all field work. Lucieer created the orthomosaic and performed the data calibration and correction procedures. Melville performed all analysis and compiled the written chapter. Lucieer and Aryal provided editorial input and advice.

We the undersigned agree with the above stated "proportion of work undertaken" for each of the above published (or submitted) peer-reviewed manuscripts contributing to this thesis:

Candidate: **Bethany Melville**

Arko Lucieer **Jagannath Aryal**

Abstract

In 2012, construction begun on Tasmania's largest irrigation project; the Midlands Water Scheme. It was determined, however, that construction and subsequent operation of the scheme could potentially negatively impact endangered lowland native grassland communities found in the region. Both the Australian government and Tasmanian State government issued a mandate for increased monitoring of these communities, however no formal program has been undertaken. Concerns were also raised about the appropriateness of the current vegetation mapping methods used in the State, as they rely on manual image interpretation techniques. Such techniques are time consuming and often produce subjective results, therefore the need for a new remotely sensed approach relying on semi-automated techniques was required. The aim of this thesis is to investigate the utility of remote sensing as a means of providing frequently updatable maps of lowland native grasslands.

The first objective of this study was to classify Tasmanian lowland native grassland communities from moderate spatial and spectral resolution satellite imagery and compare these classification results to traditional field mapping techniques. The aim of the classification was to distinguish three types of lowland grassland communities (*Poa*, *Themeda*, and grassland complex), and broader agricultural and woodland classes. An object-based image classification was undertaken on segmented Landsat ETM+ and WorldView-2 datasets using 50 random forest models trained using random subsets of reference points generated from field samples collected by the Tasmanian Land Conservancy. Validation was performed using the reciprocal points not used to train the models. Resulting average accuracies were moderately high, ranging between 55-88% for the Landsat ETM+ results, and 56-87% for the WorldView-2 result. The currently existing community map (TASVEG) was also evaluated, and found to have comparatively poorer accuracy for all classes. Analysis of Variance Results (ANOVA) indicated a significantly higher accuracy of the WorldView-2 result compared to the Landsat ETM+ result for the dry woodland and *Themeda* grassland classes, but no other statistically significant differences in classification accuracy were detected.

The results of the first study indicated the need for higher spectral resolution datasets. Therefore, the second objective of this study was to determine whether discrimination of lowland grassland communities was possible based solely on spectral properties. A field campaign was undertaken to collect data using an ASD handheld-2 spectrometer. The dataset was resampled to match the broadband resolution of Landsat OLI and WorldView-2 to compare results from narrowband and broadband approaches. Spectral signatures were classified using a random forest classifier at their full spectral resolution as well as spectrally convolved broadband equivalents to simulate coarser spectral resolutions of different sensors. ANOVA results indicated that classification accuracy for the *Themeda* class was highest when using a reduced narrowband model in which correlated bands had been removed. The analysis also indicated that significant problems in differentiating between *Danthonia* grasslands and *Themeda* grasslands at all spectral resolutions. Variable importance measures indicated strong separability in wavelengths associated with pigment decomposition, photosynthetic rate, and water content. Confusion rates and variable selections showed that differentiation between all classes was improved using high spectral resolution signatures and by grouping vegetation based on photosynthetic pathway.

Uncertainty within image segmentation can be particularly problematic in heterogeneous environments with indistinct class boundaries. Therefore, the third objective of this study was to devise a method capable of predicting class-specific optimal segmentation scale. A new method of segmentation assessment was developed that evaluated segmentation performance using a combined geometric and thematic assessment. Two trials of the index were run, the first in an urban environment with clear object boundaries, and the second in the native grassland study area in which vegetation communities intergraded significantly. Both trials indicated successful prediction of optimal segmentation scales, and the optimal segmentation scale parameters identified by the new optimisation algorithm resulted in significant improvements in class delineation and characterisation.

The fourth objective was to combine the findings of the previous three chapters in order to optimise approaches to lowland native grassland community mapping. A 15 cm, 20 band hyperspectral orthomosaic was acquired using an Unmanned Aircraft System (UAS). The spectral regions measured by the sensor corresponded to those identified previously as being of key importance. Segmentation assessment was performed using reference data acquired from transects and at additional randomly placed field plots. The UAS orthomosaic was segmented based on class-specific optimal scale predictions and classified with a random forest model. Overall, the methods outlined in the study provide a targeted approach to lowland native grassland mapping capable of providing reliable community maps at a range of scale levels. Accuracies for remotely sensed results at the landscape scale significantly improve on those from traditional mapping techniques, and therefore remote sensing is deemed to be a viable mapping approach for lowland native grasslands.

In summary, the analysis undertaken in this thesis shows that lowland native grassland communities can be accurately identified using remote sensing techniques. The results obtained here provide several key findings that illustrate the importance of data selection. Key spectral regions for differentiation of communities were identified, and can be used to improve variable and dataset selection in future analysis. This thesis contributes significantly to the broader field of grassland mapping and monitoring as it provides important case-studies proving that community level classification is possible at varying spatial scales. Finally, this thesis contributes to the field of object-based image analysis, and the prediction of optimal segmentation scale. The prediction of class-specific and scene-wide optimal segmentation scale is a novel development in the field. Additionally, the use of thematic accuracy in conjunction with spatial accuracy to determine optimal segmentation scale is a new technique developed in this thesis. Overall, the results of this thesis provide important findings that can be used to further guide conservation and targeted management of lowland native grassland communities.

Acknowledgements

Firstly, I would like to thank my supervisors, Arko Lucieer and Jagannath Aryal. The last four years of working with you both has been a wonderful experience, and I can't even begin to express my gratitude for all that you've taught me. I sincerely appreciate all the help that you've given me compiling this thesis, and the long hours that you've both put in to reading through it. Thank you for all of the opportunities you've given me, and all of the advice you've shared with me.

Secondly, I'd like to thank all of the staff and students in the spatial sciences group in the School of Land and Food. The amount of support I've received over the last few years has been absolutely phenomenal. Thank you in particular to Darren Turner, Stephen Harwin, Robert Anders, and Matt Dell for helping me with data collection and for with the more technical aspects of this project. I couldn't have done this without you. Thank you also to Iain Clarke, for putting up with me and all of my stress during the last few months of my thesis. I'd also like to thank all of the various undergraduate and graduate students I have worked with over the last few years. Thank you for reminding me of why I chose to do this, and why I love my work so much.

Thirdly, my sincerest gratitude to the Tasmanian Department of Primary Industries, Parks, Water and Environment for helping me come up with the idea for this project, and providing much needed advice and access to datasets. In particular, I would like to thank Lindsey Mitchell and Louise Gilfedder for all the help they have given me. I would also like to thank Matt Appleby from Tasmanian Land Conservancy, Kerry Bridle from the Tasmanian Institute of Agriculture, and Professor Jamie Kirkpatrick from the School of Land and Food. Thank you also to Simon Foster, for allowing us to use his property in Campbell Town as a study site, and for giving us so much insight into the Midlands, and why lowland native grasslands are so important.

On a more personal note, I would like to thank all of my beloved friends, family, and various housemates who have put up with me during this time. I could never have done this without your kindness and support. I would especially like to think my dear friend Olivia Lucas, for putting up with all my complaining and varying degrees of panic. You are a superstar.

I don't think I can even begin to describe how much of this project is owed to my long-suffering partner, James Cox. The amount of support and love you have given me over the last four years is something that I could never even begin to repay you. From driving out to the middle of nowhere on your weekends to help me take photos of grass (that you're allergic to), to reading my incoherent drafts, to talking me out of quitting, you have been there for me every step of the way, and I am so astoundingly grateful. You are my favourite. Plus, you won't have to listen to me tell you about the grass any more.

Finally, I would like to thank my parents, Rosemary and Wayne Melville. Thank you so much for inspiring me to be the best person I can possibly be, and for believing in me no matter what. I am so glad that I have parents as wonderful as you both, and I will always be so very grateful for all of the support and love you have given me. Thank you for giving me so much wonderful advice, and pushing me to finish, even when I felt like it was impossible. I could never had done this without you, thank you from the bottom of my heart.

Table of Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Tables

Chapter 2

Chapter 3

Chapter 4

Chapter 1

Introduction

1.1 Native Grasslands

With the exception of Antarctica, grasslands inhabit on every continent on Earth, providing innumerable ecological services that are often overlooked. Home to the world's prime agricultural lands, they are often viewed solely as an economic resource, with little attention being given to them as ecosystems in need of preservation. However, grasslands are as ecologically valuable as any other vegetation community type. Temperate grasslands cover over 8% of the worlds' surface, and these areas are among the most heavily altered landscapes in existence today (Species Section, 2006) .

Native grasslands are defined as 'areas of native vegetation dominated by native grasses with few or no emergent woody species' (DEWHA, 2010). In Australia, native temperate grasslands are listed as one of the most threatened communities nationwide (Harris, S., Kitchener, 2005). These grasslands form part of an important biome, found across south eastern Australia (DEWHA, 2010). The communities found in the Midlands region of Tasmania, the lowland native grasslands, are known to be unique, although they bear some similarities to temperate grasslands found in south eastern Australia (Beeton, 2006).

There have been many attempts at quantifying the extent of native grasslands in Tasmania. Gilfedder (1990) estimated that before European colonisation, there was originally 45,000 hectares of native grasslands, of which approximately 60% (27,000 hectares) remains. The Tasmanian Department of Primary Industries, Parks, Water and Environment (DPIPWE), however, estimated the extent of the grasslands to be between 85,000 hectares to 100,000 hectares in 2007 (Species Section, 2006). The estimates of decline range between 83% (Kirkpatrick et al., 1988) to 90% (Beeton, 2006). There are an estimated 1,900 patches of grassland remaining, over 80% of which are under 10 hectares in size, and 98% under 100 hectares in size (Beeton, 2006). It is believed that all grasslands that were originally on river flats have been cleared for agricultural cultivation, with the exception of a few with rocky banks or a high level of invading endemics (Fensham, R., Kirkpatrick, 1989). The observed changes in species diversity and composition indicates a reduction in the overall integrity of communities, which can lead to changes in community structure and function, finally resulting in species endangerment and extinction (Beeton, 2006). The conditions required by some lowland native grassland species are no longer naturally occurring in the Midlands, leading to subsequent endangerment and the need for artificial propagation (Gilfedder, 1990). In 2006, Beeton calculated that at the current rate of decline in the Midlands, 27% of existing communities will be lost within the next 10 years, with total extinction occurring within the next 40 years.

1.2 Midlands water scheme and potential threats to grassland communities

In April, 2012, the construction of Arthur's Pipeline, the second section of the Midlands Water Scheme, was approved by the Australian Government. This paved the way for the construction of Tasmania's largest irrigation network, designed to provide water to agricultural land located in the Midlands Region. Up to 45 000 million litres of water will be drawn from Arthurs' Lake and the South Esk River annually

and redirected through both pipelines and natural rivers to agricultural lands (DPIPWE, 2010b). Two major pipelines and a series of dams will be constructed to facilitate the redirection of water, the construction and use of which potentially poses significant risk to grassland communities in the region.

It has been estimated that 9,080 hectares of the total 21, 600 hectares of remaining lowland native grassland occur in the area designated for the Midlands Water Scheme (DPIPWE, 2010a). Key risks as identified by the Midlands Water Scheme Strategic Assessment include clearance and conversion of land, changes to plant recruitment and pollination strategies, further fragmentation of existing patches, soil nutrient change, pesticide drift, weed invasion, changes to surface and sub-surface hydrology, increased salinity, and soil compaction (DPIPWE, 2010a). The governing body for the Midlands Water Scheme has stated that every precaution will be taken in order to avoid the occurrence of any such threats, and an extensive interdepartmental monitoring scheme has been suggested in order to do so (DPIPWE, 2010b). However, concern has been raised over the lack of accurate and complete distribution mapping for grassland communities in the area, and a call has been made by the Midlands Water Scheme for this issue to be addressed further by both Tasmanian and Australian Governments.

1.3 Vegetation mapping in Tasmania

Vegetation mapping in the State of Tasmania has been an ongoing priority for both Government and non-government affiliated groups. The first digital State-wide vegetation map was released in 2001, by DPIPWE. The mapping project, eventually becoming known as TASVEG, encompassed all of mainland Tasmania, sub-Antarctic Macquarie Island, and several larger islands surrounding the State (Kitchener and Harris, 2013). A total of 162 land cover classes were identified for inclusion in the project, and formal descriptions are available in the associated TASVEG technical manual produced by Kitchener and Harris (Kitchener and Harris, 2013).

The TASVEG dataset has subsequently been through two revisions since the initial release, with the latest version, 3.0, being released in November of 2013 (DPIPWE, 2014). All land cover is mapped to a scale of 1:25,000, and derived from various imagery sources geo-referenced to an accuracy of 15 metres. The minimum mappable unit for the dataset is one hectare. The main source of imagery used to produce the TASVEG dataset consists of aerial photography (Michaels, 2006). The spatial resolution of data is typically 15 – 30 cm, with at least three spectral bands being collected over the red, green and blue regions of the electromagnetic spectrum, although some data contains a fourth near-infrared band (Michaels, 2006). In addition to the aerial imagery, satellite imagery from sensors such as RapidEye has also been employed (Michaels, 2006).

Land cover classes are identified in a number of ways, based both on manual interpretation of remotely sensed imagery (primarily aerial photography), as well as field measurements of environmental factors such as geology and topography (Michaels, 2006). Coarser scale GIS layers derived from past mapping projects are also used to incorporate environmental variables when field data is not collected for a specific region. Each of the 162 land cover types identified by the TASVEG dataset have an associated benchmark providing a comprehensive description. Benchmarks are stringently defined, and primarily used for field validation purposes (Kitchener and Harris, 2013). Forested land cover is defined based on canopy cover, understory composition, geological composition of the area, topographic variables such as slope and aspect, altitude above sea level, and mean tree height (Michaels, 2006; Kitchener and Harris, 2013). Broadly, forests fall into dry or wet sclerophyllous subtypes, rainforests, and non-eucalypt forest

or woodland. These broad definitions are then broken further into various classes typically defined by the dominant tree species (Kitchener and Harris, 2013). Non-forested communities are defined in a similar manner, with variables such as geology, environmental conditions, topography, and species composition being accounted for. In addition to providing details about the characteristic environments of various communities, vegetation benchmarks also include detailed descriptions of constituent species and the expected proportionate makeup of the community (Kitchener and Harris, 2013).

Field validation is undertaken in order to identify vegetation community types, and to provide estimates of community condition. The size of validation sites is highly variable, ranging from one hectare to greater than 100 hectares, and cover may consist of more than one target vegetation class within a single site (Michaels, 2006). Sites are then broken into a series of assessment zones of varying size and cover complexity, however, an individual zone must be created for each vegetation community identified within the bounds of the larger validation site (Michaels, 2006). Condition is assessed based on proportionate species cover and presence as defined by the relevant community benchmark, measurements of tree height and canopy closure, as well as measurements of topographic variation and soil type (Michaels, 2006). The process of field validation and assessment is well defined, however, due to the scale of the TASVEG project, there are very few validation sites, and return visits to sites are often infrequent.

The construction of Arthur's pipeline and the subsequent expansion of the Midlands Water Scheme has raised numerous concerns around the ability of the TASVEG program to provide reliable and frequently updateable maps of lowland native grassland extent. In order to meet the requirements of expanded mapping and monitoring, there is a need for scientific research into alternate mapping methods.

1.4 Remote sensing for grassland monitoring and mapping

Grassland and rangeland mapping in Australia is heavily centred on managing environmental threats, such as fire (Paltridge, 1988) drought (Mcvicar and Jupp, 1998) and salinity. These threats are typically addressed from an agricultural risk-management perspective. Studies focussing on grassland mapping from an ecological standpoint are rare within the Australian literature, with the majority of studies focussing on measurable biophysical characteristics (Graetz, 1987; Paltridge, 1988; Chladil and Nunez, 1995; Mcvicar and Jupp, 1998; Guerin *et al.*, 2017). Rangelands and grasslands within the Australian interior are used almost exclusively for grazing of sheep and cattle, and as result, research in this area is heavily focussed on agricultural applications (Graetz, 1987; Mcvicar and Jupp, 1998; Guerin *et al.*, 2017). The lack of studies focussing on grasslands as an environmental entity rather than an agricultural resource is reflected in global research trends. Internationally, the majority of research into grasslands is focussed on agricultural uses (Gao, 2007). There are however studies looking at biodiversity and environmental concerns surrounding grasslands (Guerin *et al.*, 2017). Thematic classification of grasslands at the community and species level are still uncommon within the literature, however, some studies using these approaches have proven to be successful (Wen *et al.*, 2010; Mcinnes *et al.*, 2015; Mansour *et al.*, 2016). There is a definitive need for further investigation into the feasibility of community and species level identification of grasslands, not just within Australia, but within the broader international context.

Within the State of Tasmania, there have been several attempts at mapping the extent of lowland native grassland communities, however no definitive method for generating reliable extent maps exists. Many

authors have called for more research into lowland native grasslands, and many have noted their concerns in regards to this lack of definitive community maps. Fensham and Kirkpatrick (1989) called for the undertaking of detailed botanical surveys to aid in the establishment of reserves that will protect the greatest number of species. Carter *et al.* (2003) stated that research is needed to identify remnant native grasslands found on private land, and to provide descriptions of each community type in the context of the larger bioregion. Although significant research has been undertaken in the past 40 years, there is still a great need for studies on the functional aspects, management issues, and distribution of lowland native grasslands (Carter *et al.*, 2003).

Previous attempts to map the extent and distribution of lowland native grassland communities using aerial photography and remote sensing have proved unsuccessful due to the small patch size, seasonal variation, phenology, and the difficulty of mapping at an appropriate scale (DEWHA, 2010). As many grassland species typically compose the understory in woody and grassy shrub land, it can be very difficult to map at the community level, as the highly mosaicked pattern of vegetation makes it exceedingly hard to define a meaningful minimum mappable unit. However, many of these attempts predate the development of modern spatial techniques (Beeton, 2006), and therefore, modern analysis should be tested. Accurate monitoring and mapping of vegetation communities is an area of high priority for effective environmental management and conservation. Remote sensing is extensively used to map and monitor vegetation communities around the globe (Zhang *et al.*, 2003; Reed *et al.*, 2004; Running *et al.*, 2004). The current methods used in Tasmania, however, are heavily reliant on manual interpretation of aerial photography and satellite imagery, and produce highly generalised and subjective estimates of community extent. Due to the labour intensive nature of these approaches, the period between image collection and final map generation can be quite long., which can result in outdated estimates of community extent. Additionally, field validation is limited, and often performed many years prior to the collection of the reference imagery (DPIPWE, 2014).

Given the high conservation priority of lowland native grasslands, a mapping approach based on manual interpretation of aerial and satellite imagery is deemed to be inappropriate. A mapping approach for species management and assessment needs to provide an accurate and repeatable methodology that can identify and assess communities at an appropriate spatial scale. Remote sensing is therefore proposed as an alternative method, with the aim of reducing the issues associated with the current mapping approach. Remote sensing is a well-established tool in the field of vegetation mapping, and has been used extensively for both monitoring and mapping of various grasslands and rangelands (Tucker, Justice and Prince, 1986; Williamson, 1992; Pickup, Basin and Chewings, 1994; Tieszen *et al.*, 1997; Guo *et al.*, 2004; Marsett *et al.*, 2006; Rango *et al.*, 2009; Foody and Dash, 2010; Laliberte *et al.*, 2011). For example, recent applications include large-scale estimates of grassland biomass in China (Becker *et al.*, 2007), evaluating species richness based on seasonal variation (Wang *et al.*, 2016), and mapping grassland degradation (Mansour *et al.*, 2016).

1.5. Object-based classification

In the last decade, Object-based Image Analysis (OBIA) has come to the forefront of remote sensing analysis techniques. Object-based classification operates on the concept of image objects, as opposed to single pixels. These objects relate to one another through an inbuilt hierarchical structure, and can be modelled with both shape and topological features, in addition to traditional spectral identifiers (Bruzzone *et al*, 2006). Each object is composed of a group of pixels, each with similar spectral

properties (Cleve *et al.*, 2008). Object-based analysis attempts to describe the relationships between image objects in terms of their characteristics, such as spectral information, texture, shape, area, and scale (Repaka *et al.*, 2004). By integrating textural, contextual and spectral information into the classification process, OBIA is able to achieve significantly higher classification accuracy than pixel-based classifiers in many situations (Yan *et al.*, 2006; Blaschke *et al.*, 2014). Objects created using this technique are more spectrally homogenous within their individual regions than they are between their neighbours. They have clearly defined boundaries, are distinct from each other, and are representative of real-world entities (Yu *et al.*, 2006).

One of the key advantages of OBIA as a method for landscape mapping is its ability to incorporate contextual features derived from multiple data sources (Baatz and Schäpe, 2000). OBIA approaches to environmental analysis are capable of deriving ecologically meaningful estimates of habitat distribution when the communities in question are not easily separable on the basis of spectral information alone (Bock *et al.*, 2005). This advantage makes OBIA particularly applicable in heterogeneous landscapes where community distribution is often determined by spatial characteristics rather than spectral characteristics alone (Gibbes *et al.*, 2010). Given the dynamic nature and typically high level of fragmentation observed in remnant native grassland patches, these characteristics of OBIA approaches are highly applicable.

As image segmentation forms the basis of all OBIA techniques, there has been a great deal of research in recent years aimed at creating a method of determining optimal segmentation scale. Optimal segmentation scale is generally assumed to be the scale at which image segments correspond perfectly to their associated reference objects (Möller *et al.*, 2007; Weidner, 2008; Clinton *et al.*, 2010; Whiteside *et al.*, 2014). Many methods have been developed, primarily using geometric agreement metrics to quantify spatial agreement between segments and reference objects (Clinton *et al.*, 2010). There have, however, been very few attempts to quantify accuracy in terms of thematic attributes or agreement. This remains a current gap in the knowledge base.

Due to the highly fragmented and varying size of remnant lowland native grassland patches, it is incredibly important that the appropriate scale of analysis is correctly determined. Within OBIA approaches, it is well known that final classification accuracy is highly dependent on the accuracy of the underlying initial image segmentation (Blaschke *et al.*, 2014). However, current methods for assessing segmentation accuracy typically focus only on geometric properties, and correspondence between the relative spatial placement of image segments and reference objects (Neubert *et al.*, 2006; Clinton *et al.*, 2010; Whiteside *et al.*, 2014). This causes issues for cases such as lowland native grasslands, where the potential benefits, such as improved performance in extremely heterogeneous environments, of using OBIA approaches are great. The requirement of clearly defined reference objects with discrete boundaries is problematic in natural environments such as this, and therefore the existing approaches for predicting optimal segmentation scale will not necessarily be applicable or appropriate. Therefore, there is a need for the establishment of a new method of segmentation assessment that can determine optimal segmentation scale using non-geometric object properties. The development of such a metric would allow for the reduction of uncertainty around the determination of the optimal scale of analysis which has proven to be a significant barrier in the establishment of remote monitoring schemes for these communities in the past.

1.6 Research Objectives

The overall aim of this project was to create a new remote sensing approach to community mapping of lowland native grasslands in the Midlands region of Tasmania. The approach should be capable of producing accurate, reliable, and easily updateable maps of community extent. In order to achieve this goal, appropriate datasets that can be used in such an approach needed to be identified, and the optimal spectral and spatial properties required for accurate lowland native grassland mapping needed to be determined. In order to successfully achieve these aims, however, the overarching issue of determining optimal segmentation scale within OBIA-based approaches had to also be addressed.

Objective 1

To determine the suitability of moderate to high resolution multispectral satellite imagery for identification and mapping of lowland native grassland communities. Additionally, the ability of pre-existing coarse resolution vegetation maps to act as reference datasets was assessed. Multispectral satellite datasets are easily acquired and frequently captured, and many vegetation maps have been created for the region through manual digitisation of aerial photography. The application of these datasets within a new mapping framework is potentially capable of providing broad-scale analysis of communities that can be repeated with greater frequency than current manual and field-based methods. By utilising existing vegetation maps as reference data sources, the need for expensive field work can potentially be reduced.

Objective 2

To identify the spectral properties of remnant communities in order to determine the best approach for spectral differentiation. The utility of narrowband and broadband approaches needs to be tested in order to determine the required spectral resolution for accurate community identification. Knowledge about the optimal spectral resolution for separating dominant grassland species will aid in the selection of future data sources.

Objective 3

To develop a method for predicting optimal segmentation scale for multiple thematic classes that considers both geometric and thematic attributes of image segments and reference objects. The limited success of previous mapping attempts for lowland native grasslands can be attributed primarily to inappropriate scales of analysis; therefore there is a need to determine the optimal image segmentation scale in order to accurately match the spatial scale of remnant grassland community patches with the scale of analysis.

Objective 4

To test the feasibility of unmanned aircraft systems (UAS) as a platform for collecting data at the optimal spatial and spectral resolutions identified by previous objectives. This objective aims to test the optimal data characteristics identified under previous objectives, and bring them together within a single case study.

1.7. Thesis Structure

This thesis consists of four core chapters (chapters 2-5) aimed at meeting the requirements of a thesis by publication. Chapter 2 addresses the first study objective, and provides an assessment of readily available

multispectral datasets. Chapter 3 contains a detailed spectral analysis of lowland native grassland communities, and determines the spectral resolution required for successful discrimination between lowland community types, in order to address Objective 2. Chapter 4 considers Objective 3, presenting a method for predicting optimal segmentation scale for multiple thematic classes, and two case-studies demonstrating proof of concept. Chapter 5 presents the results of a high-spatial and spectral resolution approach for mapping lowland native grasslands using imagery acquired with an unmanned aircraft system (UAS). It employs the methodologies developed in previous chapters, and presents a final optimised community mapping approach, as stated in Objective 4. Chapter 6 presents the final conclusions of the work presented in chapters 2-5.

Chapter 2

Object-based Random Forest Classification of Landsat ETM+ and WorldView-2 satellite imagery for Mapping Lowland Native Grassland Communities in the Tasmanian Midlands Region

Abstract

This chapter presents a top-down approach for identifying and mapping three types of lowland native grassland communities found in the Tasmanian Midlands region. The purpose of this study was to assess the effectiveness of classifying multispectral satellite datasets as an alternative to manual digitisation of vegetation community extents. Two datasets were acquired in the spring of 2010; a Landsat ETM+ scene, and a WorldView-2 scene. A dataset collected by the Tasmanian Land Conservancy (TLC) was used as a reference dataset for classification training and validation. A copy of the Tasmanian Vegetation map, known as TASVEG, was acquired for the study site. A series of reference points were generated from the TLC dataset with a minimum spacing of 30 m between points, as this corresponds to the spatial resolution of the coarsest satellite dataset (Landsat ETM+). Both satellite images were segmented using eCognition, and object-based texture measures calculated. Additionally, a 25 m digital elevation model (DEM) was acquired, and various topographic measures applied.

Training and validation data were randomly subset from the full set of reference points at a ratio of 66% training to 33% validation within each class. A total of five land cover classes were identified; three lowland grassland communities, a broad dry woodland class and an agriculture class. The reference dataset was randomly split into training and validation a total of 50 times in order to employ a *k*-folds cross-validation approach to classification training and validation using a random forest model. For each subset of points, a random forest model was trained for the Landsat ETM+ dataset and the WorldView-2 dataset. These models were then used to classify each segmented satellite dataset 50 times. Validation was performed using the reciprocal points from the random subset that weren't used to train the model. Final training and classification accuracies were reported as means per class across the 50 classifications performed on each satellite dataset. The TASVEG map for the study site was also evaluated against the range of validation points. Analysis of Variance (ANOVA) was undertaken in order to determine whether observed differences in classification accuracy differed between the two types of satellite imagery, as well as between classification trials.

Results indicate good performance for the majority of classes within both satellite dataset classification results, although poor classification accuracy (~55%) was achieved for the grassland complex class in both sets of results. It was found that class accuracy only differed significantly between the two classification results for the dry woodland and *Themeda* classes, with the WorldView-2 dataset showing higher mean classification accuracies. When compared to the TASVEG vegetation map, all classes exhibited significantly improved classification based on both datasets. Overall, the classification results show improvement in community delineation when compared to the TASVEG community map, showing that remote sensing is a viable method for lowland native grassland mapping.

1 Introduction

Remote sensing analysis of grasslands has typically been focussed on agricultural applications, due to their high economic value. Studies therefore tend to focus on measurement of grassland biophysical characteristics, rather than on community classification or identification. The most common research topics that use multispectral remote sensing approaches focus on quantifying grassland biomass, cover, or degradation (Gao, 2007). Multispectral sensor platforms commonly used to undertake such analysis include Landsat (Goetz, 1997; Price *et al.*, 2002; Marsett *et al.*, 2006; DPIPWE, 2010a; Zerger *et al.*, 2011; Yang *et al.*, 2012), SPOT (Williamson, 1992; Goetz, 1997; He and Guo, 2006), and MODIS (Dubinin *et al.*, 2010; Wen *et al.*, 2010; Xu *et al.*, 2013; Sim and Abdullah, 2014; Jin *et al.*, 2015). Estimation of such characteristics is typically based on the use of vegetation indices, in particular the Normalised Difference Vegetation Index (NDVI) (Goetz, 1997; Tieszen *et al.*, 1997; Ji and Peters, 2003; Seaquist *et al.*, 2003; Dilley *et al.*, 2004; Dusseux *et al.*, 2014). Index values are usually statistically related to field measurements of the target characteristic, and the derived relationship used to extrapolate across a larger study area. Such approaches allow for broad-scale analysis of grassland attributes, and are particularly useful in circumstances in which monitoring is required over large areas. Thematic classification-based approaches with the aim of differentiating between types of grasslands are uncommon within the literature. The most frequently found method of differentiating between communities is on the basis of photosynthetic pathway, most notably between C_3 and C_4 groups (Tieszen *et al.*, 1997; Foody and Dash, 2010; Yang *et al.*, 2012). Despite the rarity of these approaches, some authors have had success in species or community level discrimination using thematic classification approaches on multispectral satellite datasets (Wen *et al.*, 2010; Mcinnes *et al.*, 2015; Mansour *et al.*, 2016), although the use of multispectral sensors mounted on Unmanned Aerial Systems (UAS) is now becoming increasingly common due to their higher spatial resolution (Laliberte *et al.*, 2007; Rango *et al.*, 2009; Laliberte *et al.*, 2011; Lu *et al.*, 2016). Multispectral satellite datasets, however, have some inherent advantages over UAS approaches for broad-scale community mapping in that; 1) they have larger spatial coverage within a single scene, 2) legacy operations mean that datasets may be available over several decades, 3) they often cover a broader spectral range than consumer-grade sensors commonly used in UAS research and 4) datasets are often freely available for public use.

The current approaches to vegetation mapping employed within the State of Tasmania are largely based on manual interpretation of imagery, such as aerial photography. Although these approaches can provide a good estimate of vegetation composition and extent, there are limitations to the applicability of such approaches. A major concern is the lack of repeatability and the subjectivity inherent with manual digitisation. Extent predictions will often vary between analysts, and even between repeat digitisations of a scene by the same individual. Therefore, with the advent of modern remote sensing techniques, there exists an opportunity to update such methods using a more repeatable and objective approach. The existence of high quality, readily available multispectral datasets is a valuable resource that should be investigated. In order to meet the mandate of increased monitoring of lowland native grassland communities, there is first a need for the generation of accurate community maps. Whether classification of multispectral satellite datasets can provide such maps is therefore an important area of research, as updated monitoring cannot be undertaken without improved estimates of community extent.

1.1 Aims and objectives

The aim of this study was to develop an object-based image analysis classification approach using multispectral satellite datasets that is capable of identifying lowland native grassland communities with

a higher degree of accuracy than current mapping approaches. Of primary concern was the identification of the spatial scale at which vegetation communities become most separable, therefore two satellite datasets of varying spatial resolution were trialled as possible input data. By employing such an approach to community mapping, it was hoped that the accuracy of maps can be improved, and the time frame between consecutive maps greatly reduced. This provided greater ability to both monitor and identify remnant lowland native grassland community patches, as well as an improved ability to detect changes in community condition across the larger Midlands area. Such an approach to mapping will potentially greatly reduce the expenditure of both financial and human resources required to produce community maps, as it will remove the need for extensive manual digitisation.

2. Methods

2.1 Study Site

The study site is situated on a private farm located between the Ross and Campbell Town in the Tasmanian Midlands region. A subset of the property covering 16.9 km^2 was been selected. The land is predominantly used for grazing of sheep and cattle, although some commercial crops are also grown. There are a number of lowland native grassland communities found on the property, in addition to introduced sown pasture species. The lowland native grassland communities on the site are well maintained and managed, and are among some of the highest quality remaining communities in the State. Fertilisation and irrigation are both used throughout the site to increase productivity, but some pastures are left unmanaged. *Poa labillardierei* communities are primarily found growing along the banks of the creek that runs through the centre of the site and in the far north-west. *Themeda triandra* communities are found primarily in the north-eastern corner. The western edge is bound by a large hill populated by dry eucalypt woodland species. The geology is a complex of basic igneous, sedimentary arenaceous, and sedimentary argillaceous rocks (DPIPWE, 2010b) resulting in a mixture of sodosol, tenosol, and chromosol soils. The average annual precipitation ranges from 375 mm to 625 mm. Figure 2.1 shows the location of the study site.

A total of 12 land cover classes were identified in the TASVEG dataset, while the TLC data identified a total of 14. Due to the large number of identified classes, and in order to keep the classes used consistent between the TASVEG and TLC datasets, the decision was made to create a series of broader land cover classes. As the primary target of investigation are lowland native grassland communities, the three major community types present in both datasets, lowland native grassland complex, lowland native *Poa labillardierei* grasslands, and lowland native *Themeda triandra* grasslands, were retained. The three original forest and woodland classes were merged to form the new Dry Woodland class. Finally, remaining land cover classes consisting of wetlands and non-native vegetation, such as agriculture, silviculture, weed infestations, and regenerated cleared land were merged to form a broad agricultural class.

The Table 2.1 lists the range of land cover classes originally identified, the dataset in which they appear, the area covered by the class in each dataset, a brief description of the community based on the TASVEG benchmarks, the dominant cover species, and the final simplified class assignment. Additionally, Figure 2.2 shows the extent and vegetation communities of the TASVEG and TLC data.

2.2.2. Classification input data

As one of the primary goals of this study is to identify the spatial resolution of input data required to produce an accurate vegetation community classification, imagery from two different multispectral satellite platforms was employed. By classifying two sets of satellite imagery independently, the impact of spatial resolution can be investigated. Landsat 7 ETM+ imagery was obtained for October 19th, 2010. 2010 was selected as the study year due to its clear weather during the target spring growing season, and its proximity to the period in which the TASVEG 3.0 dataset was in production. The Landsat ETM+ sensor has 8 spectral bands, 6 of which cover the blue to near infrared portions of the visible spectrum (490-1100nm). The remaining two bands consist of a thermal band and a panchromatic band. All bands were included in the study except for the thermal (band 6) and panchromatic (band 8) due to differing spatial resolutions. Each pixel represents an area of 30 m by 30 m. The imagery was radiometrically calibrated using the ENVI 5.1 Landsat Calibration utility, and atmospherically corrected using the Quick Atmospheric Correction (QuAC) module. Upon acquisition from the United States Geological Survey (USGS), the imagery had already been geometrically corrected. WorldView-2 imagery was obtained for November 17th, 2010. WorldView-2 has 8 spectral bands, ranging from ultraviolet/blue to near-infrared (400-900 nm) with a spatial resolution of 2.0 m. The imagery was atmospherically corrected using the QuAC module in ENVI 5.1 (Exelis Visual Solutions, 2013). The same atmospheric and aerosol models were selected as for the Landsat scene As the TASVEG classification scheme is based on both visual appearances of communities and environmental variables such as topography and geology, a 25 m LiDAR derived Digital Elevation Model (DEM) with five meter positional accuracy was obtained from DPIPWE through Land Information Systems Tasmania (LIST). From the DEM model, slope, and aspect were derived using ArcGIS 10.3 (Environmental Systems Research Institute, 2014). All input data were projected using the Universal Transverse Mercator projection, with a zone of MGA 55S. The datum used was the Geocentric Datum of Australia, 1994.

2.2.3 Image segmentation

Image segmentation was performed in eCognition version 9 (Trimble Navigation, 2014) using the multiresolution segmentation algorithm (Baatz and Schäpe, 2000). Segmentation was performed independently for the Landsat ETM+ and WorldView-2 data. For both datasets, the full set of atmospherically corrected multispectral bands, in addition to the DEM, aspect and slope models was used for segmentation. For the WorldView-2 segmentation, the topographic data was included after resampling to a 2 m spatial resolution using a cubic convolution algorithm. A scale factor of 15 was used for the Landsat ETM+ segmentation, which produced a total of 575 image objects within the image. A scale factor of 375 was used for the WorldView-2 data, creating 1,524 objects within the image. Segmentation scale factor refers to the input variable in the multiresolution image segmentation that controls for object size. Scale factors are unit-less, and represent the maximum internal variability that an object can encompass before the object must be split. Image segments are grown from a randomly allocated seed until the values within the object boundaries exceed the scale factor threshold, at which point a new segment is created. Larger scale factors generally result in larger image objects, whilst smaller segmentation scale factors result in smaller objects. The scale factors used here were determined using visual interpretation of segmentation results. For both segmentations the shape and compactness factors were assigned values of 0.1 and 0.7 respectively. The shape factor determines the proportionate weighting between colour and shape criteria in the segmentation, while the compactness factor controls the compactness of resulting objects (Baatz and Schäpe, 2000). For this criteria, values closer to one allow for more compaction. The compactness factor used here was determined by visual inspection.

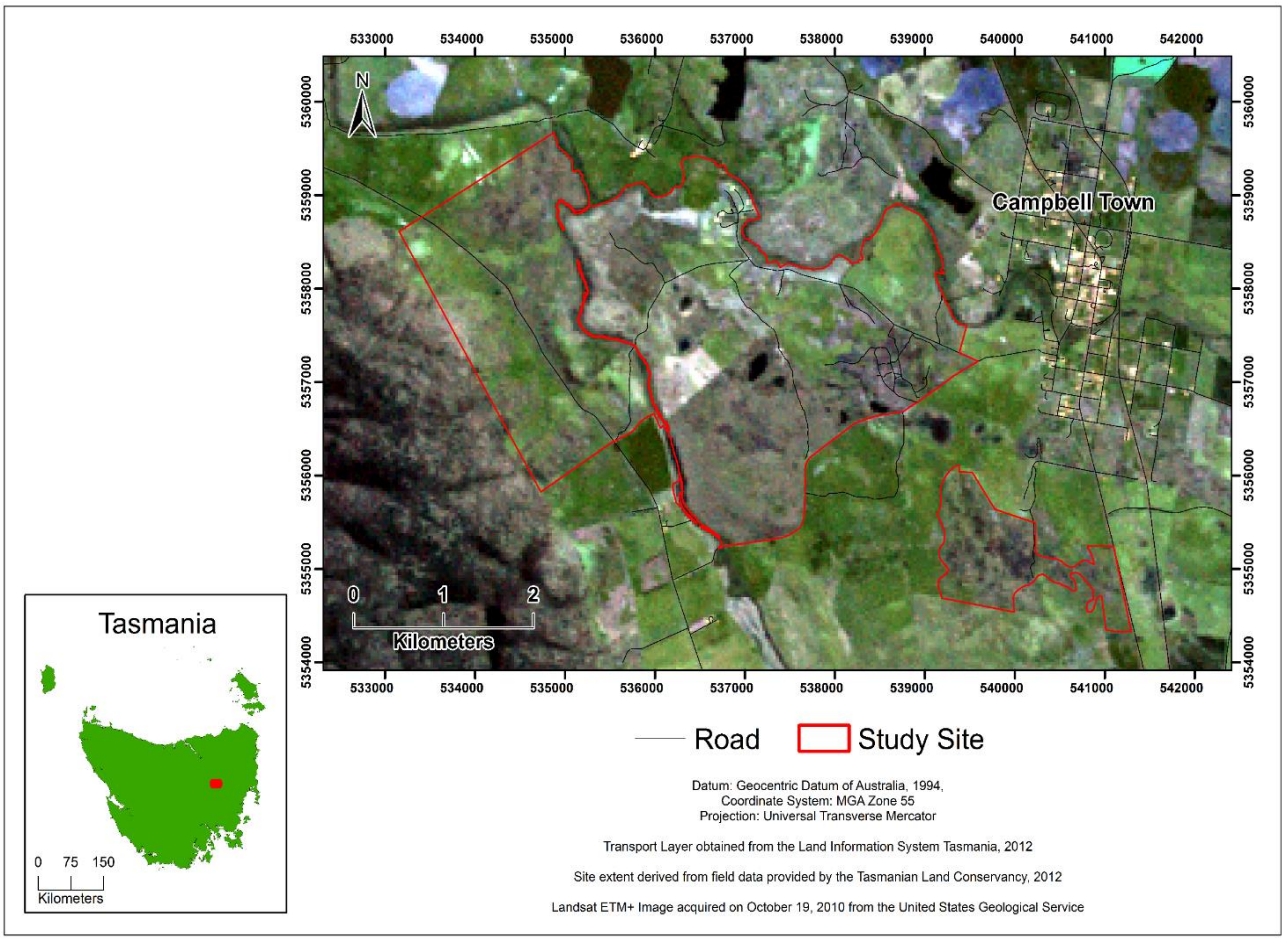

Figure 2.1: Location of the study site as shown relative to the town of Campbell Town. The extent of the study site is derived from TLC field data obtained in March, 2012. The base satellite data is a 30 m Landsat ETM+ image acquired on the 19th of October, 2010, loaded as an RGB colour composite.

Additionally, within eCognition, a series of object-based texture measures were derived. Object-based texture measure differ from pixel-based measures in that instead of employing a square kernel with user-defined dimensions, texture measures are calculated using object boundaries as the area of analysis. Texture measures were included as additional variables as it has been shown that such measures are strongly related to vegetation height and structure (Wood *et al.*, 2012)**,** which can be difficult to quantify using spectral measures alone. As variations in land management and vegetation structure are key differentiating factors between native and non-native pastures, these measures were included.

Table 2.1: Identified vegetation communities from both TASVEG 3.0 data and TLC data for the study site. TASVEG codes are listed in addition to their corresponding area within each dataset. Additionally, community definitions as per the TASVEG 3.0 benchmarks and descriptions are listed as well as the dominant cover species. The final simplified class assignment is listed in the final column.

TASVEG Code	Data Set	Area (km^2) (TASVEG)	Area (km^2) (TLC)	Description	Dominant Species/ Land use	Final Class
ASF	TLC		0.215	Succulent saline herbland	*Sarcocornia quinqueflora*	Agriculture
AHL	TASVEG	0.048		Lacustrine herbland	*Selliera radicans*	Agriculture
ASS	TLC		0.095	Freshwater aquatic sedgeland and rushland	*Baumea* spp., *Carex* spp., *Cyperus* spp., *Eleocharis* spp., *Gahnia* spp.	Agriculture
AWU	TASVEG/TLC	0.276	0.025	Wetland (Undifferentiatied)	NA	Agriculture
DPO	TLC		0.377	*Eucalpytus pauciflora* forest and woodland not on dolerite	*Eucalyptus pauciflora*	Dry Woodland
DVG	TASVEG/TLC	1.444	1.017	*Eucalyptus viminalis* grassy forest and woodland	*Eucalyptus viminalis*	Dry Woodland
FAG	TASVEG/TLC	5.191	4.221	Modified agricultural land	Improved pastures and croplands	Agriculture
FPL	TLC		0.001	Plantations for silviculture	*Eucalyptus nitens, Eucalyptus globulus, Pinus radiata*	Agriculture
FPU	TASVEG	0.040		Unverified plantations for silviculture	*Eucalyptus nitens, Eucalyptus globulus, Pinus radiata*	Agriculture
FRG	TASVEG/TLC	0.092	0.005	Regenerating cleared land	Abandoned farmland	Agriculture
FWU	TASVEG/TLC	0.401	0.360	Weed infestation	*Marrubium vulgare, Lycium ferocissimum, Salix fragilis*	Agriculture
GCL	TASVEG/TLC	1.813	1.236	Lowland grassland complex	*Rhytosperma* spp., *Austrostipa* spp.	Grassland Complex
GPL	TASVEG/TLC	1.781	2.957	Lowland *Poa labillardierei* grassland	*Poa labillardierei*	*Poa* grassland
GSL	TASVEG/TLC	0.045	0.035	Lowland grassy sedgeland	*Lepidosperma* spp., *Lamandra longifolia* spp.	Agriculture
GTL	TASVEG/TLC	2.260	2.365	Lowland *Themeda triandra* grassland	*Themeda triandra*	*Themeda* grassland
NBA	TASVEG/TLC	3.662	4.028	*Bursaria- Acacia* woodland	*Bursaria spinosa, Dodonaea viscosa, Acacia dealbata, Acacia mearnsii, Acacia melanoxylon, Acacia verticillata*	Dry Woodland

As texture measures are calculated on a single-band basis, the bands in which the various land cover classes showed the most variability and differentiation were identified for each dataset. Variability and differentiation were assessed by overlaying the TLC data on the Landsat ETM+ and Worldview-2 datasets respectively, and calculating Region of Interest (ROI) statistics in ENVI 5.1 based on the TLC class boundaries. For the Landsat ETM+ data, the fourth (850 nm) and fifth (1650 nm) bands were identified as having the largest differences between classes, and for the WorldView-2 data the sixth (730 nm) and eighth (880 nm) bands were identified. These bands were then used as the basis for object-based texture calculations. Within eCognition, four co-occurrence texture measures were calculated for each of the identified bands. The measures of homogeneity, contrast, entropy and angular second moment (energy) were calculated in all directions. These measures were selected as they were identified by Harlick et al. (1973) to provide a good overall assessment of image texture when used together.

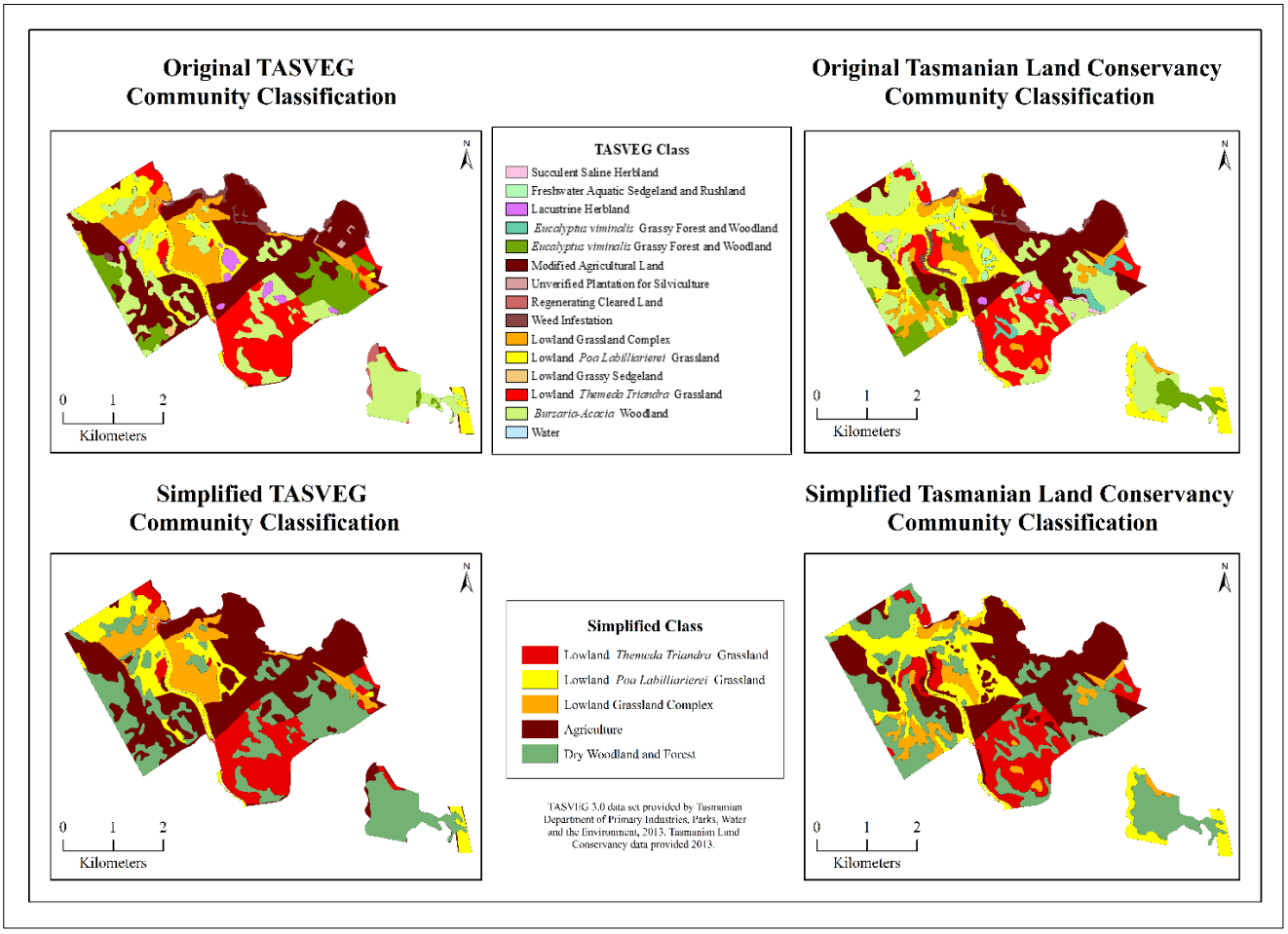

Figure 2.2 : Original TASVEG and TLC community classifications, in addition to the simplified class scheme used in the analysis. High priority lowland native grassland classes (Lowland *Poa* grassland, lowland *Themeda* grassland, and lowland native grassland complex) were retained in the simplified scheme, while all other classes were merged into either the dry woodland, or agriculture.

2.4. Random Forest (RF) training and classification

Classification of the imagery was undertaken using an RF Classifier. RF is an ensemble classifier utilising multiple decision trees to make predictions of class allocation (Duro *et al.*, 2012). An RF is trained from a series of input variables, of which one is the target, using bootstrapping (Naidoo *et al.*, 2012). Trees are built through recursive binary partitioning, which divides the data into homogeneous

regions called nodes. A random subset of the input variables is selected and tried at each node of the tree (Duro *et al.*, 2012), with the decision to split based on the data values within the random subset of variables (Watts and Lawrence, 2008). One of the key advantages of RF classification is the calculation of variable importance measures for the classification as a whole, and for individual classes (Cutler *et al.*, 2007; Dorigo *et al.*, 2012). The relative importance of variables is calculated by determining the permuting variables values across the range of the dataset, and calculating the relative change in classification accuracy of the model based on the permuted and original results. The best model is then identified with a weighted voting process aimed at selecting the most significant predictor variables from the input (Naidoo *et al.*, 2012), however these estimates have been known to show bias when the predictor variables are highly correlated.

For training and validation purposes, 3,443 points were generated within the boundary of the TLC map using the random point tool in ArcMap 10.3. The minimum distance between points was set to the spatial resolution of the Landsat ETM+ dataset in order to avoid sampling individual pixels more than once. The final sample size of 3,443 points is based on the maximum number of points that can be randomly generated within the bounds of the study site based on the minimum distance threshold of 30 m. Class labels for both validation and training were derived from the TLC dataset. To ensure that any observed differences in classification accuracy between the two input datasets is due to the effect of differing sensor properties, the generated point set was randomly split 50 times into training and validation datasets at a ratio of 66% training to 33% validation. This process was undertaken to ensure that the sampling regime and placement of training and validation points did not contribute to final differences in classification accuracy. Within the random forest algorithm itself, 33% of training points are set aside to be used for cross-validation for each tree of the RF model, meaning that only 66% of the input data is used to train each tree. This can be problematic when training datasets are small, or when the number of observations for specific classes are low, as the full statistical distribution of a class is not necessarily used. Each random subset of points was used to train, classify, and validate a random forest model for each of the two satellite datasets.

For both datasets, a series of RF models were trained using the randomForest package (Liaw and Wiener, 2002) in R (R Core Team, 2015). Training was undertaken using spectral, textural and topographic variables. The points used to train subsequent models were identical between the datasets, but the input variables were derived only from one sensor source. This resulted in a total of 50 RF models for each dataset based on random configurations of training points selected from the full point dataset. All models were trained using the same input parameters, with the number of trees set to 1000, and the number of variables (mtry) equal to four. Four was selected as $\sqrt{m}$ has been established as the optimal value for mtry (Breiman, 2001; Naidoo *et al.*, 2012) where m is equal to the total number of input variables (Breiman, 2001). From the training models, variable importance measures as well as out-of-bag (OOB) error estimates for each class and the overall model were obtained. OOB errors are commonly derived from RF training models, as they provide an estimate of model accuracy. Error rates for each class are predicted using bootstrapping (Breiman, 2001). OOB errors for each of the training subsets were averaged to find the mean training accuracy per-class for both datasets. In addition, variable importance measures were also derived and averaged across the 50 training results for each class and the overall result.

After training was complete, the segmented Landsat ETM+ and WorldView-2 images were classified 50 times using the output RF models generated from the training data subsets using the randomForest package in R. The input data for each classification was identical for all trials within a dataset, with only the RF model used to assign the class labels differing between subsequent classifications. This process was undertaken independently for both the WorldView-2 and Landsat ETM+ datasets, with both datasets being classified based only on their own segmented values.

Validation was undertaken for each output classification model using the remaining 33% of data points not used to train the RF model. Generalised classification maps for both datasets were produced by identifying the majority class across all classification results. In order to determine whether image classification using satellite-derived datasets is capable of creating more accurate vegetation community maps than the previously existing TASVEG dataset, the TASVEG dataset was resampled to the simplified class system, and assessed against the same subsets of validation points used to assess the classification outputs from both datasets. To determine the degree of uncertainty in the classification results, frequency of class assignment statistics were calculated for both the WorldView-2 and Landsat ETM+ results. For each segment within a dataset, the number of times it was classified as a given class was determined. Segments that have been identified as multiple classes across the 50 classification trials indicate classification uncertainty.

Once all the classification results and the TASVEG dataset had been validated, an analysis of variance (ANOVA) was undertaken in order to determine whether the observed differences in classification accuracy between the results obtained for the three datasets were statistically significant. Significance was tested for using a factorial ANOVA design to determine the significance of the input dataset, the subset of training and validation points used to generate and validation the RF model, and vegetation class. Once the significant effects had been determined, data was split by class in order to determine whether mean classification accuracy differed within classes based on the dataset and configuration of training and validation points. Tukeys' Post-Hoc comparisons were undertaken to determine which means differed significantly within classes based on the input dataset, and to identify any individual training and validation subsets with statistically different outcomes.

3 Results

3.1 RF training OOB accuracy estimates and variable importance measures

The Landsat ETM+ training results averaged across all 50 models are displayed in Table 2.2. The mean overall accuracy was 76.56% ± 0.81%. Class accuracies are generally high, although the grassland complex class has significantly poorer accuracies with an average score of 54.59% ± 4.08%. The agricultural class is the strongest performer, while the dry woodland, *Poa* grassland and *Themeda* grassland classes all obtained similarly high accuracies.

Table 2.3 shows the training accuracies obtained from the 50 training trials undertaken for the WorldView-2 dataset. The mean accuracy across all trails was equal to 77.96% ± 0.82%. Class specific accuracies are very similar to the results obtained from the Landsat ETM+ model. All three native grassland classes have slightly higher accuracies in the WorldView-2 result, but the increase is small. The dry woodland class also exhibits a minor increase in accuracy, while the agriculture class shows a decrease. The standard deviations obtained from each class are also similar to those obtained from the

Landsat ETM+ results, with the exception of the grassland complex in which the deviation has decreased.

Table 2.2: Confusion matrix for OOB accuracies averaged across the 50 randomly generated training subsets for the Landsat ETM+ RF training model. Confusion matrix values represent average pixel counts, while the overall accuracy is shown as a percentage. Standard deviation values represent one standard deviation in percentage points from the mean.

	Dry Woodland	**Agriculture**	**Grassland Complex**	***Poa* Grassland**	***Themeda* Grassland**	**User's Accuracy**	**Standard Deviation**
Dry Woodland	618.56	41.22	31.44	51.22	57.56	77.32	± 1.41
Agriculture	37.14	580.2	6.52	28.84	11.3	87.38	± 0.88
Grassland Complex	39.28	11.58	99.9	20.02	12.22	54.59	± 4.08
***Poa* Grassland**	53.62	38.28	21.3	306.46	9.34	71.44	± 1.72
***Themeda* Grassland**	60.9	9.92	9.24	12.82	236.12	71.77	± 2.33

Table 2.3: Confusion matrix and RF training OOB accuracy estimates for WorldView-2 model. Confusion matrix value show pixel counts attributed to each class, while accuracies and standard deviations are supplied as percentages

	Dry Woodland	**Agriculture**	**Grassland Complex**	***Poa* Grassland**	***Themeda* Grassland**	**User's Accuracy**	**Standard Deviation**
Dry Woodland	629.16	42.32	21.3	51.66	55.56	78.65	±1.24
Agriculture	45.08	576.52	5.72	31.3	5.38	86.83	± 0.99
Grassland Complex	35.72	20.18	102.08	15.88	9.14	55.78	± 3.01
***Poa* Grassland**	51.86	40.38	13.12	313.98	9.66	73.19	± 1.80
***Themeda* Grassland**	52.08	5.84	5.64	12.2	253.24	76.97	± 2.09

Figures 2.4 and 2.5 show the class-specific average variable importance measures for both the Landsat ETM+ and WorldView-2 training datasets. Results are averaged across the total set of 50 RF models per dataset, with standard deviation additionally shown for each variable. Within the Landsat ETM+ data, the variables with high importance scores are consistent across the majority of classes. Aspect, near infrared reflectance (840 nm), the green reflectance (560 nm), slope and band 5 (1,650 nm) homogeneity are identified as key variables in all classes, and for the training models overall. Poorly performing variables include band 4 entropy, band 5 entropy, and band 5 second moment. Importance levels across the WorldView-2 result are much more variable between classes, although elevation is consistently the strongest performing variable. The most important spectral band is band 6 (730 nm), which has high importance scores for all classes. Texture measure importance is variable between classes. Band 7 contrast appears to be the most consistently important texture measure when the results are taken as a whole, although several classes have different texture measure identified as being the most important.

3.2 RF Classification Results

Figure 2.3 displays the classification results for the Landsat ETM+ and WorldView-2 datasets. The final class of each segment is determined by taking the majority class assignment across the 50 classifications. Agreement between the majority classification results obtained for the two datasets is high. Differences between the two results are primarily a factor of different segmentation scales and spatial resolutions, with the WorldView-2 result identifying several small patches of classes not found in the Landsat-ETM+ result. Boundary delineation of classes also appears to differ between the two results, with the Landsat ETM+ result showing more distinct changes between classes than the WorldView-2 result. Figure 2.6 shows the frequency of various class assignments for each object. Values are tabulated by finding the number of times each object is assigned to a given class across all classification trials, and then converting the count to a percentage.

For the Landsat ETM+ result, confusion between classes is much more prevalent in the four native vegetation types than in the non-native agriculture class. Objects with a high level of uncertainty are most frequently found in the boundary areas between two or more classes, particularly for the dry woodland class. The most inconsistent class is the grassland complex, which is expected given the poor accuracies obtained during training. The WorldView-2 classification frequency counts show slightly lower levels of uncertainty than the Landsat ETM+ data. Objects with low frequency counts are observed in the boundary zones between classes. The most variable class in terms of classification frequency of objects is the dry woodland class, with many transitional areas showing moderate degrees of classification uncertainty. The *Themeda* grassland class in this case also exhibits some changes in frequency that were not observed in the Landsat ETM+.

3.3 RF Classification and TASVEG Validation

The overall classification accuracies for the Landsat ETM+ RF classifications are shown in table 2.4. Confusion matrix counts and per-class percentages are averaged across the results of the 50 classifications. The average accuracy across all classes and all trials is 76.72% ± 0.83%. The highest class accuracy was obtained for the agricultural class, and the lowest for the grassland complex, with standard deviation showing a similar pattern. Accuracies and standard deviations are similar to those obtained from the training trials, with accuracies typically varying by less than 1%.

Table 2.4: Final classification accuracy for Landsat ETM+ RF classification. Standard deviation and accuracy statistics are provided as percentages, while confusion matrix values show average pixel counts:

	Dry Woodland	**Agriculture**	**Grassland Complex**	***Poa* Grassland**	***Themeda* Grassland**	**User's Accuracy**	**Standard Deviation**
Dry Woodland	316.98	21.88	15.92	27.48	28.74	77.12	± 2.03
Agriculture	19.04	298.58	3.12	14.42	5.92	87.54	± 1.97
Grassland Complex	19.72	6.36	50.98	9.34	6.6	54.82	± 5.61
***Poa* Grassland**	27.66	20.2	10.22	157.76	4.08	71.74	± 3.50
***Themeda* Grassland**	30.14	5.12	4.62	6.5	121.62	72.39	± 3.60

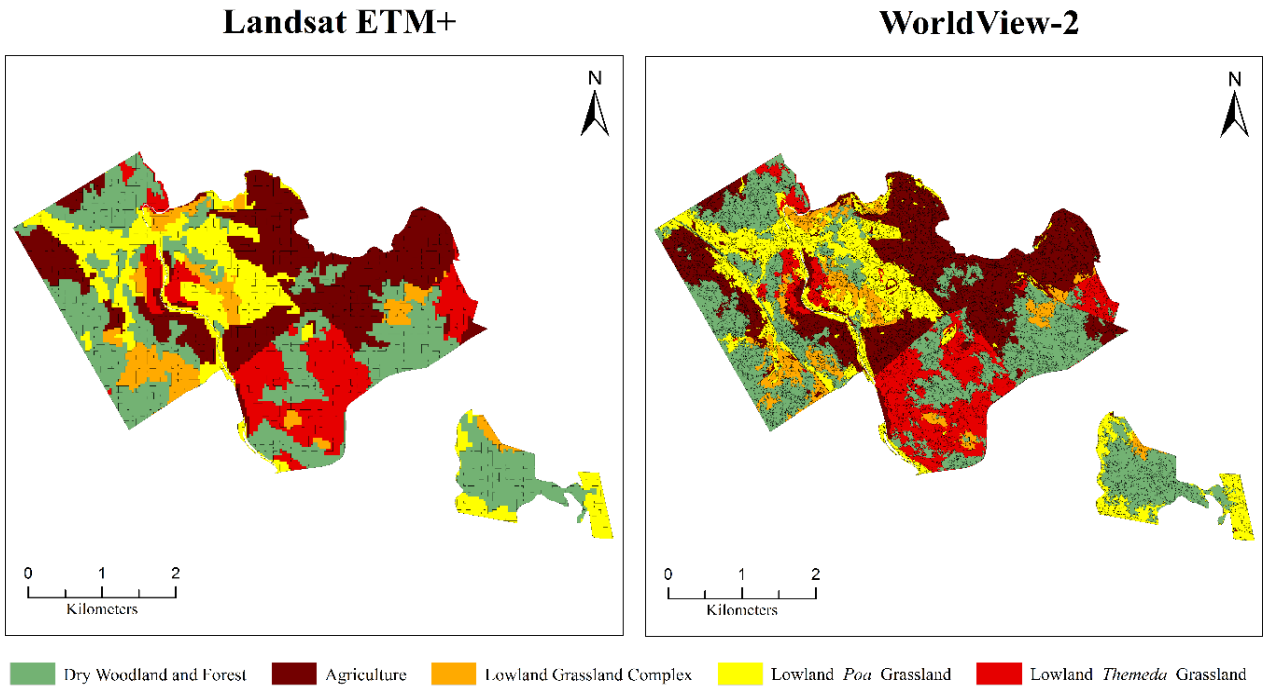

Figure 2.3: RF majority classification results for the Landsat ETM+ and World-View datasets. Class assignment for each object is derived from the majority object assignment across all 50 trials. Boundaries of image segments are delineated with black lines.

Final mean classification accuracies obtained for the 50 WorldView-2 RF classifications are shown in table 2.5. Confusion matrix counts are provided as a mean pixel count, while accuracies are given as a percentage. Standard deviation values represent percentage point variation from the mean. The final mean accuracy across all results is equal to 78.27% ± 0.95%. As observed in the previous results for the Landsat ETM+ model, classification accuracies are very similar to the training accuracies, with all classes varying by approximately 1% each. Standard deviations are slightly higher in the validation results, with an increase of around 2% observed for the majority of classes from the training results.

Table 2.5: Final accuracy for WorldView-2 RF classification trials. Confusion matrix values are given as pixel counts, while accuracy and standard deviation are given as percentages

	Dry Woodland	**Agriculture**	**Grassland Complex**	***Poa* Grassland**	***Themeda* Grassland**	**User's Accuracy**	**Standard Deviation**
Dry Woodland	325.12	21.94	11.06	26.7	26.18	79.10	±1.97
Agriculture	22.98	298.28	3.02	14.42	2.38	87.47	±2.28
Grassland Complex	17.58	10.64	52.32	7.4	5.06	56.26	±4.87
***Poa* Grassland**	28.1	19.38	7.22	160.38	4.84	72.9	±3.45
***Themeda* Grassland**	26.32	2.9	3.02	6.78	128.98	76.77	±4.09

Landsat ETM+ RF Variable Importance Measures

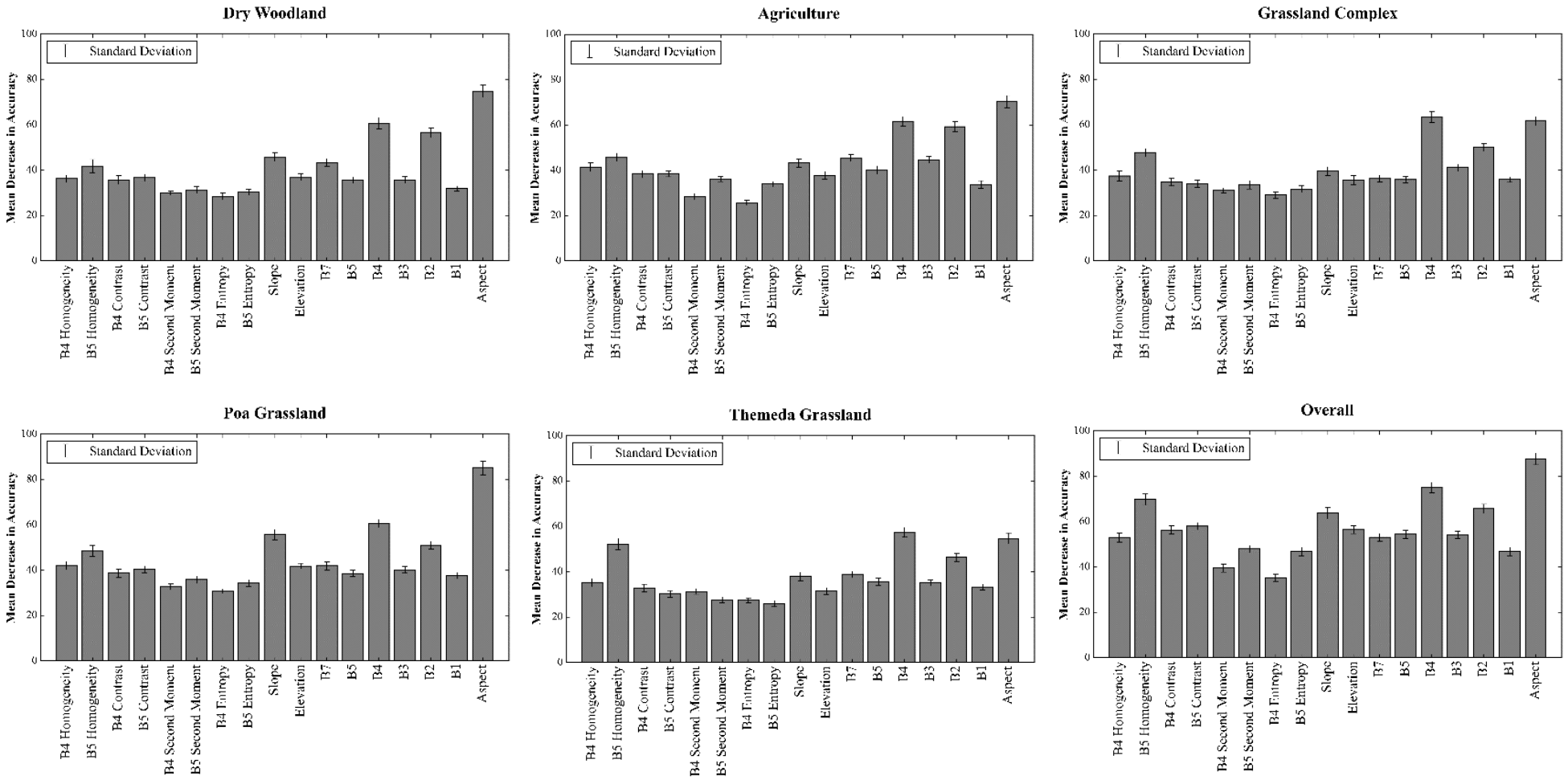

Figure 2.4 RF variable importance measures obtained from Landsat ETM+ training averaged across 50 trials. Importance values are determined by randomly permuting variable values across the dataset and calculating the resulting decrease in classification accuracy between the original and permuted result. High importance scores indicate variables containing important information for class discrimination and accurate classification

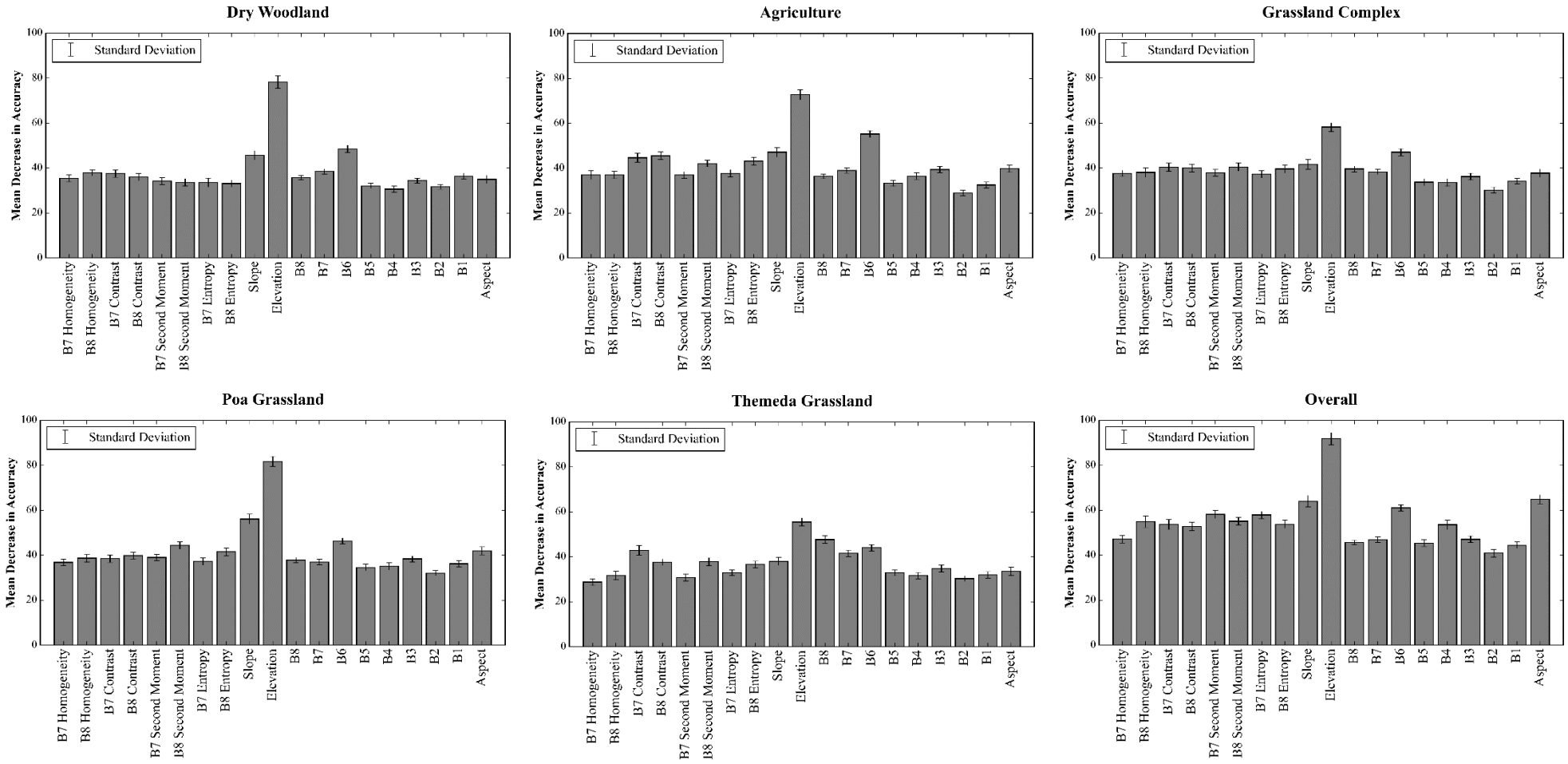

Figure 2.5: Mean RF variable importance measures obtained for the WorldView-2 training results. As can be seen, elevation is consistently important for all classes.

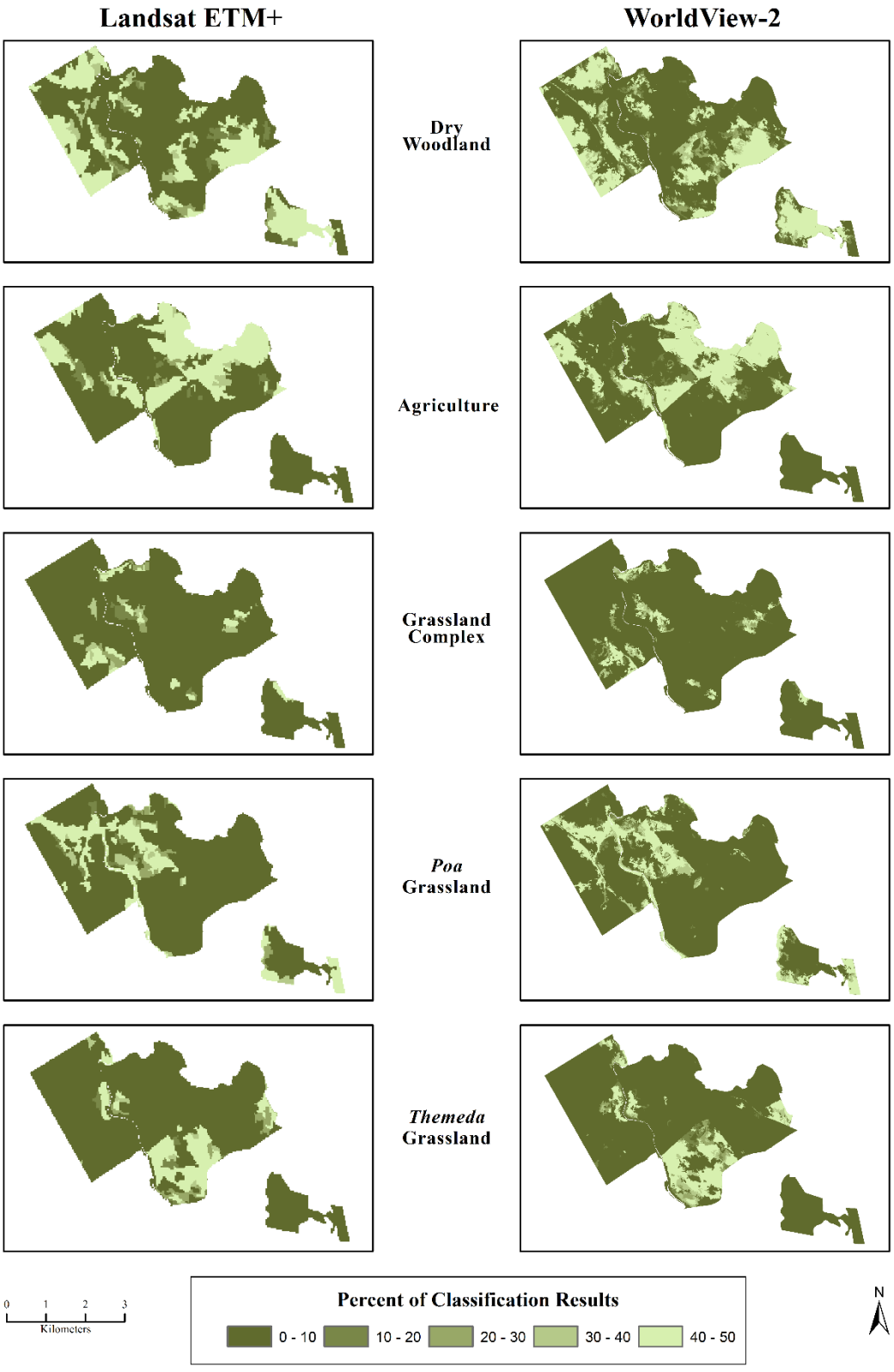

Figure 2.6: Landsat ETM+ and WorldView-2 per-class classification frequency counts. Colour gradations show the percentage of classification results in which each segment was identified as belonging to a certain class.

The accuracy assessment results for the TASVEG validation process are displayed in table 2.6. Accuracies were determined using the same validation point sets used to assess the Landsat ETM+ and WorldView-2 classification results. The overall mean accuracy of the dataset is 65.09% with a standard deviation of ± 0.86%. Accuracies for the dry woodland and agriculture classes are similar to those obtained from the classification results, however, accuracies for the three native grassland classes are much lower. Both the *Poa* grassland and grassland complex have extremely poor mean accuracies, at 31.76% ±2.44% and 26.30% ± 3.44% respectively.

Table 2.6: Confusion matrix for TASVEG dataset. Assessment of accuracy was undertaken using the same subsets of validation points used to assess the Landsat ETM+ and WorldView-2 classification results. Confusion matrix values represent mean pixel counts across trials, while accuracy and standard deviation are in percentage points.

	Dry Woodland	Agriculture	Grassland Complex	Poa Grassland	Themeda Grassland	User's Accuracy	Standard Deviation
Dry Woodland	304.04	33.44	26.34	24.86	22.32	73.98	±1.58
Agriculture	16.9	288.74	12.58	15.28	7.58	84.65	±1.57
Grassland Complex	8	46.4	24.46	3.68	10.46	26.30	±3.44
Poa Grassland	27.28	58.72	59.36	69.84	4.72	31.76	±2.44
Themeda Grassland	19.38	10.54	12.76	9.78	115.54	68.77	±2.72

3.4 ANOVA and Tukeys' Post-Hoc Comparisons

The results of the ANOVA model comparing the effects of the dataset, vegetation class and the training and validation subset used for model generation and validation indicated the presence of significant effects on classification accuracy at $p \leq 0.05$ for the dataset and vegetation class. The effect of the training and validation subset used was determined to be non-significant at $p \leq 0.05$.

Class specific ANOVA results are summarised in table 2.7. The results indicate that the satellite dataset is a significant determining factor of class accuracy for all classes, while subset of training and validation points was non-significant for all classes. The post-hoc mean comparisons indicate that the TASVEG dataset has significantly poorer accuracy than both classification results when compared to the TLC validation data. There is a significant difference between the results of the Landsat ETM+ and WorldView-2 results in only three cases; the dry woodland class, the *Themeda* grassland, and the overall classification accuracy. In all three of these cases, the difference detected is a higher mean classification accuracy in the WorldView-2 result over the Landsat ETM+ result. Additionally, in these three cases, the Landsat ETM+ result has a significantly higher classification accuracy than the TASVEG dataset. For the remaining three cases of agriculture, grassland complex, and *Poa* grassland, the observed increase in classification accuracy for the WorldView-2 results over the Landsat

ETM+ results is non-significant. All three instances however have significantly poorer accuracy in TASVEG than in the Landsat ETM+ and WorldView-2 classification results.

Table 2.7: Results of ANOVA analysis showing significant and non-significant variables in determining classification accuracy for all vegetation classes. The two right hand columns indicate the variable levels with and without significant differences in mean classification accuracy for each class.

	ANOVA Result (p≤0.05)		**Tukeys' Post-Hoc Comparisons (Confidence Interval=95%)**	
Class	**Significant Factors**	**Non-Significant Factors**	**Significant Different**	**No Significant Difference**
Dry Woodland	Dataset	Subset	TASVEG, WorldView-2, Landsat ETM+	N/A
Agriculture	Dataset	Subset	TASVEG	WorldView-2, Landsat ETM+
Grassland Complex	Dataset	Subset	TASVEG	WorldView-2, Landsat ETM+
***Poa* Grassland**	Dataset	Subset	TASVEG	WorldView-2, Landsat ETM+
***Themeda* Grassland**	Dataset	Subset	TASVEG, WorldView-2, Landsat ETM+	N/A
Overall	Dataset	Subset	TASVEG, WorldView-2, Landsat ETM+	N/A

4 Discussion

4.1 Training and Classification Performance

Training and classification accuracies, as well as standard deviations, are very similar within the Landsat ETM+ and WorldView-2 results. Such similarities between mean training and classification accuracy indicate that the training process is robust, and that there is consistency in the statistical distributions of the training and validation subsets. For all classes, observed differences in mean training and classification accuracy are less than 1% for both. Additionally, classes have low standard deviations, which indicates high consistency in the class accuracies achieved for all classifications. In general, large differences in the degree of accuracy between training and validation results indicates potential sampling bias and the inability of the training data to encompass the variability of the full dataset, thereby resulting in poor classification accuracy relative to training accuracy. Excessively high training accuracies compared to validation accuracies can also indicate potential over-fitting of random forest models. Overall, the consistency of results within and between models in this study indicates that the use of such a sampling scheme can be an effective. Additionally, the results show that the method can be effective in cases where the number of potential training points is low.

The variable importance measures also show a high consistency, with all variables exhibiting low standard deviations around the mean accuracy scores for each class. For both sets of RF

models, all variables show some degree of importance for each class. Across both datasets, topographic variables are consistently identified as having a high importance for all classes. Topographic variables are a key component of many class benchmarks within the TASVEG system, especially for lowland native grassland communities. The inclusion of topographic variables in the approach was based on their use in the community benchmarks, and a desire to use similar variables to those used in the production of the original TASVEG dataset. Given this, the high variable importance scores assigned to these variables are not unexpected. Spectrally, the greatest differentiation appears to be associated with the near-infrared portion of the spectrum, with high importance being associated with Landsat ETM+ band 4 (770-900 nm), and WorldView-2 band 6 (705-745 nm). These selections indicate that differences in photosynthetic activity may be key potential differentiators between grassland classes. Variable importance for the texture measures is far more variable within the WorldView-2 model than the Landsat ETM+ model, where band 5 (630-690 nm) homogeneity is consistently identified as being important for a number of classes in the WorldView-2 result. The similar importance values attributed to the spectral and textural variables across the two datasets indicates strong differences in vegetation structure between community types, which again is expected given variations in canopy types between introduced and native grass species found throughout the study site. Detailed hyperspectral investigation of key spectral regions, and the biophysical properties of vegetation associated with these regions would be an important source of information for selecting appropriate sensors and improved use of texture measures.

4.2 Class performance and confusion rates

For both classification results, mean accuracies are similar between the two sensors. Class-specific mean accuracies are very similar between the two results, with only two classes, dry woodland and *Themeda* grassland showing significantly higher accuracies in the WorldView-2 result compared to the Landsat ETM+ results. In addition, overall mean accuracy was also shown to be significantly higher in the WorldView-2 result.

The poorest performing class in both classification results was the grassland complex class, with mean classification accuracies of 54.82% for Landsat ETM+ and 56.26% for WorldView-2. This class also has the highest standard deviation of all classes in both sets of results. The poor performance and high variability observed for this class is likely due to the broad definition of the class in terms of species composition and dominance. The grassland complex is the most loosely defined of the lowland native grassland classes, and serves as a generalised class accounting for any native grassland not dominated by *Poa labillardierei* or *Themeda triandra* (Kitchener and Harris, 2013). Typically, the community is dominated by *Danthonia* species, but the formal community benchmark states that other grassland species may be dominant or co-dominant (Kitchener and Harris, 2013).

In the Landsat ETM+ results, misclassification of the grassland complex class is primarily due to confusion with the woodland class, and secondarily with the *Poa* class. Confusion between native grassland classes and the woodland class is expected given the high degree of community intergrading. In the WorldView-2 result, confusion is again greatest with the woodland class, but the second highest rate of confusion is with the agriculture class. Confusion

with other native grassland classes is very low, and likely due to the increased number of bands in the red edge region of the spectrum, which are better able to detect varying phenology between communities. The high level of confusion with the agriculture class is likely due to the smaller segmentation scale employed for the WorldView-2 dataset. Due to the fact that the grassland complex and agricultural classes often share class borders, any potential misplacement of segment boundaries is likely to manifest as misclassification between the two classes.

The best performing class in both sets of results is the agricultural class. This class has the advantage of being the most distinct in terms of its spectral and textural properties. The majority of species incorporated in this class are introduced, and have planophile canopies. All other classes are also composed of native vegetation, which exhibits typical sclerophyllous adaptations and morphology, while the introduced agricultural species do not. These differences result in the agricultural class having distinctive reflectance properties associated with high photosynthetic rate, greenness, and water content. Additionally, the agricultural land in the study site is managed much more intensively than the native areas. Such management practices include the use of irrigation, fertilisation, and the sowing of both crop and pasture species. These practices result in distinctive textural properties, such as clear row marks from sowing, that clearly differentiate them from more native areas where growth is less constrained.

The dry woodland class is the second highest performing class for both results. ANOVA results indicate that there is a significant improvement in classification accuracy for this class in the WorldView-2 data over Landsat ETM+ results. Confusion values are similar between the datasets. The exception to this, however, is the grassland complex class, which has a reduction in confusion for the WorldView-2 result. The definition for the woodland class is quite broad, as it covers three forested land cover classes as originally identified in the TLC data. Two of the original classes, DPO and DVG, are *Eucalyptus* dominated woodland variants, with low levels of floristic diversity in their understories (Kitchener and Harris, 2013). The remaining class, NBA, is an open woodland dominated by *Acacia* or *Bursaria* species over a dense grassy understorey, typically of *Themeda triandra* or *Danthonia sp.* (Kitchener and Harris, 2013). The classification frequency counts observed for the woodland class in figure 2.6 indicate that the areas with higher rates of misclassification are associated primarily with the known distribution of the NBA class, as shown in figure 2.2. The overlap in dominant species between the NBA class and the native grassland classes, coupled with the sparseness of the tree cover associated with this type of woodland, is the source of significant confusion within the classification results for both datasets.

The *Poa* grassland class shows good performance in both sets of results, with no significant difference in accuracy between the two datasets. Confusion for this class is primarily with the dry woodland and agricultural classes. Misclassification with the woodland class is due to reasons discussed previously, such as overlapping constituent species between classes. Both the agricultural and *Poa* classes have similar phenological cycles, as species are primarily C_3 species. At the time of acquisition, both classes were entering a period of senescence over the warmer summer months, which may have resulted in confusion due to lower rates of

photosynthetic activity. Additionally, some of the area covered by the *Poa* class is actively managed through the use of irrigation and fertiliser, though not as extensively as the designated management practices in both classes may be an additional source of confusion due to increased variance in class spectral properties. There is very little observable confusion with the *Themeda* class which occurs almost exclusively in unmanaged areas of the property.

The final native grassland class, *Themeda*, has the strongest classification performance of the three lowland native grassland types, and additionally has significantly improved performance in the WorldView-2 results. Confusion for this class is primarily with the woodland class, as per reasons discussed previously. For the Landsat ETM+ result, confusion with the remaining three classes occurs at similar levels for each, while in the WorldView-2 result, confusion with the grassland complex and agriculture classes is greatly reduced. The main differentiating factor for the *Themeda* class is its varying phenology and photosynthetic pathway as it is a C_4 species. *Themeda triandra* also exhibits a very characteristic red colouration, which, coupled with the increasing growth rate of the species at the time of data acquisition, may be a reason for its improved performance in the WorldView-2 result, given the increased number of bands in the red and red-edge regions of the spectrum.

4.3 Performance of remotely sensed results in comparison to TASVEG

The confusion between classes in the Landsat ETM+ and WorldView-2 datasets differ substantially from those observed in the TASVEG dataset. For all classes, both sets of classification results were identified within the ANOVA outputs as having significantly higher classification accuracies. Final accuracies for the agriculture and dry woodland classes are similar to those obtained from the classification results, differing by only a few percentage points. The performance of the three grassland classes, however, is poor in comparison, especially for the *Poa* and grassland complex classes.

In terms of class confusion, the dry woodland class is confused more with the agriculture and grassland complex classes in the TASVEG dataset, but confusion with the *Themeda* and *Poa* grassland classes is approximately the same, although there is a slight decrease in the confusion for *Themeda*. As can be seen in figure 2.2, the extent of the simplified woodland class in the TASVEG and TLC datasets is very similar, explaining the high accuracy of the class. Wooded areas are the most readily identified in the scene by visual interpretation, which would facilitate accurate digitisation. The agricultural class also has very similar extents in the two vector datasets, and exhibits consistent levels of confusion in both the TASVEG and classification results. The extent of the agricultural class is also readily apparent in the imagery, given the presence of fences, which cause hard transitional boundaries between areas of agriculture and native vegetation, which again facilitates accurate digitisation.

Grassland complex was the poorest performing class, achieving an accuracy of only 26% in the TASVEG accuracy assessment. Figure 2.2 shows widely differing predictions of extent between the TASVEG and TLC datasets, with TASVEG predicting extensive areas of grassland complex in the north-west of the scene. The majority of grassland complex areas within the TLC data have been misclassified as agriculture in TASVEG. There is also very

little overlap with grassland complex areas and the *Poa* and dry woodlands classes, but a slight increase in confusion with *Themeda* grassland is observed. Areas associated with the grassland complex class in the TLC dataset are areas with scattered woodland cover that are also lightly managed through the application of irrigation and fertilisers. The occurrence of these practices may have led to improper identification of the area as agriculture, given the increased greenness and vigour of the vegetation as a result of active land management.

The *Poa* class, like the grassland complex class, also exhibited extremely poor accuracies in the TASVEG result. Figure 2.2 shows an extensive over-prediction of grassland complex, especially in the northern half of the scene. The TLC data identifies the majority of this area as being *Poa* grassland, which accounts for the confusion between classes. Both of these classes have very similar dominant species, which have similar phenological cycles and canopy structure. These similarities in species cover are the likely reason for such extensive misclassification in the TASVEG digitisation result.

The *Themeda* class is again the best performing of the grassland classes, although there is a decrease in the overall accuracy as achieved in the classification results. Confusion between the grassland complex and agriculture classes has increased significantly. The large patch of *Themeda* located in the bottom portion of the scene is properly identified in the TASVEG result, however, several of the smaller areas identified by the TLC data have been missed, or significantly reduced in size, resulting in the observed increased confusion.

4.4 Dataset resolutions and segmentation parameter selection

Classification accuracy in an OBIA environment is primarily determined by the accuracy of the segmentation scale (Neubert *et al.*, 2006; Weidner, 2006). This can prove to be quite problematic, as there is currently no universally accepted metric for the assessment of segmentation accuracy. In order for a classification to be accurate, the segmented objects must resemble real-world features (Whiteside *et al.*, 2011). Several metrics have been proposed that provide an estimate of agreement between segments and real-world objects (for example (Lucieer and Stein, 2002)), however, there is no metric that assesses the robustness of a segmentation in relation to thematic classes. Spatial agreement between scene and ground objects is most certainly an integral consideration in the assessment and determination of an appropriate segmentation scale, however, there also needs to be an agreement between the established class characteristics and the characteristics exhibited by individual scene objects. The thematic accuracy of an image segmentation can be defined as the degree to which image segments represent the thematic attributes of the classes to which they will ultimately be assigned. Although the optimal segmentation scales for an image in terms of thematic accuracy and geometric accuracy often co-occur, this is not always the case. It is important that both forms of accuracy are optimised within a segmentation in order for it to provide the best possible representation of real-world objects. It is important to remember, however, that measures of segmentation accuracy and spatial agreement are not measures of classification accuracy (Clinton *et al.*, 2008). This is of particular concern in highly heterogeneous environments, such as native grasslands, that may benefit from the generalisation associated with image segmentation. In many cases however, the establishment of hard class boundaries

can be difficult due to strong community intergradation. Therefore, thematic agreement between segments and reference objects is an important consideration in the assessment of segmentation quality and accuracy in this context.

High spectral resolution allows for more detailed spectral signatures to be acquired for each class. Similar classes are easier to distinguish from each other in high spectral resolution datasets with many narrow spectral bands, given the potentially subtle differences in spectral signatures between grassland classes. The WorldView-2 sensor has eight spectral bands, four of which cover the portion of the spectrum between the red and infrared (700 -1,100 nm). Landsat ETM+ has six spectral bands, three of which cover the red and infrared portion of the spectrum. The inclusion of an extra band in the area of 700-1100 nm allows for potentially greater differentiation between classes in the WorldView-2 imagery as opposed to the Landsat ETM+ imagery given an appropriate scale of analysis. However, Landsat ETM+ has two spectral bands in the near-infrared to shortwave infrared portion of the spectrum, which potentially provides valuable information for the identification of vegetation.

Grasslands have been proven to be exceptionally responsive to seasonal changes, and are typically found in areas that exhibit strong seasonal variations in key environmental factors (Tieszen *et al.*, 1997). Therefore, accounting for seasonal variation is an important factor in the identification and differentiation of grassland communities. Tiezen *et al.* (Foody and Dash, 2010) as well as Foody and Dash (Tieszen *et al.*, 1997) have highlighted the key role that seasonal difference between communities with similar physiological properties and habitat distributions can play in correct identification. The inclusion of temporal data derived from Landsat imagery was tested for this study, but the results were unsatisfactory given the spatial scale of the imagery. Future work could benefit greatly from the inclusion of temporal variables.

5 Conclusions

This chapter outlines an updated approach for mapping lowland native grasslands in the Tasmanian Midlands region using remote sensing methods. This approach provides significant improvements in classification accuracy for all vegetation communities over the TASVEG dataset. Additionally, the methods outlined in this chapter are capable of being regularly repeated, which is an important consideration given the Australian Governments' mandate for increased community mapping and monitoring.

Two satellite datasets with differing resolutions were trialled as potential classification inputs, Landsat ETM+ and WorldView-2. Additionally, a 25 m DEM was acquired, and various topographic variables were derived for inclusion in the classification models. Object-based texture measures were calculated based on key spectral bands for both datasets. Training and validation data were derived from a pre-existing data source collected by the TLC. A random subset of training points was generated, ensuring a minimum point spacing of 30 m (the coarsest sensor spatial resolution employed in the study) in order to avoid oversampling of individual pixels. Training and validation data were randomly split from the original reference

point dataset using a ratio of 66% training to 33% validation a total of 50 times in order to employ a *k*-folds cross-validation approach to model training and validation. This approach was selected in order to reduce sampling bias, and ensure that all possible data points were used to train and validate the resulting models. Classification was undertaken using a random forest classifier, with each subset of training points used to train two separate models; one derived from the Landsat ETM+ data, and one from the WorldView-2. Classification was then performed on segmented versions of both datasets, with the reciprocal reference points not used to train the model used for validation. Class accuracies were averaged across the 50 classification results for each dataset, and classification frequency counts for each class tabulated.

Overall, both models showed good results for all classes, with class specific accuracies ranging between 54-87% for the Landsat ETM+ classifications, and 56-87% for the WorldVIew-2 classifications. The performance of the grassland complex class was significantly lower than for other classes, averaging 54% for Landsat ETM+ and 56% for WorldView-2. Classification and training accuracies for all classes across both models showed a high degree of consistency relative to each other, and standard deviations for all classes were low. This indicates that there is no bias introduced into the classification and assessment process as a result of training and validation point selection. Additionally, the TASVEG dataset was validated against all 50 validation subsets, and accuracies compared to the classification results. ANOVA indicated that for all classes, resulting accuracies were significantly higher in both sets of classification result than for TASVEG. The analysis also indicated significant improvements in *Themeda* grassland, dry woodland and overall accuracies based on the WorldView-2 dataset over the Landsat ETM+.

In conclusion, this study meets the demands of a remotely sensed classification approach that can cover larger areas. This approach can be used to map the spatial extent of grassland communities at an increased temporal resolution given the availability of cloud-free satellite imagery However, due to the use of coarse resolution training data, the approach may not be able to identify fine-scale changes in community distribution, and is best suited to the generation of similarly low spatial resolution results. Despite this however, achieved classification accuracies across both sets of results indicate that multispectral satellite datasets are capable of providing accurate extent predictions for lowland native grassland communities in the Midlands region.

6. Thesis Context

This chapter provides a top-down approach to lowland native grassland community mapping, and outlines an object-based image classification approach capable of providing updated and improved maps of community extent. The results obtained indicate that remote sensing is a viable option for community mapping. However, several issues have been identified, and overall classification accuracies are relatively modest. Several areas in need of improvement have been noted, such as the need for higher spatial resolution datasets, and also to identify key spectral properties associated with different grassland communities. Issues surrounding

segmentation scale are also of particular concern, especially given the heterogeneous nature of the communities in question. Future chapters will aim to address these issues, and determine the best practices required for accurate community delineation.

Chapter 3

High spectral resolution analysis of lowland native grassland communities based on field image spectroscopy

Abstract

This chapter presents the findings of a study designed to determine the optimal spectral resolution required to differentiate between lowland native grassland communities. A field campaign was undertaken to collect spectral signatures of different community types using a portable field spectroradiometer with a 3 nm spectral resolution. A series of four datasets were created from the field spectra; a full high spectral resolution narrowband set utilising all available spectral bands collected by the spectroradiometer, a reduced high spectral resolution narrowband set in which highly correlated bands were removed, a convoluted Landsat 8 set in which the original high spectral resolution field data were resampled to match the spectral bands of Landsat 8, and a simulated WorldView-2 dataset in which the high spectral resolution bands were resampled to the spectral resolution of WorldView-2.

Two groups of land cover classes were created, a geneneralised three-class model, and a more specific four-class model. The four-class model consisted of saltpan, *Wilsonia rotundifolia/Sellieria radicans, Danthonia trenuior* grassland representing the grassland complex class identified in TASVEG, and *Themeda triandra* grassland, while the three-class model consisted of saltpan, C_3 vegetation (made up of the merged *Wilsonia* and *Danthonia* classes), and *Themeda triandra* grassland. Classification was undertaken using a random forest approach. Training and validation datasets were subset using *k*-fold cross-validation, in which the total set of field points were randomly subset into training and validation datasets 30 times. Classification was then performed using the generated training models 30 times for each dataset using both a three-class and four class configuration. Mean training and classification accuracies showed a high level of consistency across results within datasets, indicating that the configuration of training and validation points did not introduce significant bias. Achieved classification accuracies were good for all classes in the three-class tests, ranging from 85-93% across classes. Accuracies for the four-class tests were similar to the three-class results for the saltpan and *Themeda* classes, however the performance of the *Danthonia* and *Wilsonia* classes were poor, averaging ~80% for *Wilsonia* and 56% for *Danthonia.*

Analysis of Variance (ANOVA) results indicated that for the three-class tests, there was no significant difference in classification accuracy between the four datasets. For the four-class tests, ANOVA indicated that there was significant improvement in the *Themeda* result when using the reduced high spectral resolution dataset. The results of the ANOVA tests undertaken between classes showed that the merged C_3 class exhibited significantly higher classification accuracy than the split *Wilsonia* and *Danthonia* classes, and that the four-class *Themeda* result was significantly improved over the three-class result. The overall conclusion of the study is

that high spectral resolution datasets are required for optimal classification of *Themeda triandra* grasslands, but for all other trialled classes, broadband approaches are sufficient.

1 Introduction

The purpose of this chapter is to determine if the use of high spectral resolution (HSR) narrowband datasets can provide an improvement in class separability between lowland native grassland communities. HSR data has been shown to have many advantages over broadband multispectral data, particularly for grassland communities bearing significant similarities to one another. For example, Mutanga and Skidmore (Mutanga and Skidmore, 2004) found that narrowband indices generated from hyperspectral data were able to provide better estimates of grassland biomass than broadband equivalents. Many narrowband indices and isolated regions of the electromagnetic spectrum have been found to be strongly correlated with grassland properties. For example, significant relationships have been found between levels of dry and wet biomass present in grasslands and spectral reflectance in the region of the spectrum between 350 nm and 450 nm (Tucker, 1977). Other authors have also found relationships between the red and NIR portions of the spectrum and differences in grassland biomass (Mutanga and Skidmore, 2004; Cho and Skidmore, 2009; Gianelle and Guastella, 2016). Although the majority of publications using HSR narrowband datasets for grassland analysis focus on prediction of biomass rather than community classification, the findings of such research still provide a valuable framework from which to build classification approaches. In fact, the findings of such previous research would indicate a strong potential for community differentiation based on biophysical and biochemical parameters such as biomass, pigment levels, and water content (Tucker, 1977). In addition, consistent issues with multispectral analysis of complex vegetation communities have been identified, highlighting the need for more detailed HSR approaches (Roth *et al.*, 2015)

Narrowband spectral analysis of communities has also been postulated as a potential source of reliable validation data for broad-band multispectral approaches (Roth *et al.*, 2015). The caveat of such approaches, however, is that there is a definite need for the spatial scale of analysis to be carefully considered and optimised to produce reliable estimates of vegetation parameters (Rahman *et al.*, 2003; Roth *et al.*, 2015). This chapter presents the results of a detailed spectral analysis performed using HSR narrowband field measurements in order to determine optimal spectral regions for class differentiation and subsequent classification. These measurements have been carried out in such a way as to remove spatial resolution from consideration, in order to only assess the influence of spectral variables on classification outcomes for lowland native grasslands.

1.1 Aims and Objectives

The aim of this study was to determine if lowland native grassland communities can be accurately classified solely based on their spectral properties. Key regions of spectral separability between classes were determined based on spectral signatures collected with a narrowband handheld field spectroradiometer. Field observations were then resampled to resemble broadband satellite observations in order to determine if accurate classification of

grassland communities is possible using a broadband spectral analysis approach. This will in turn guide future sensor selection and measurement protocols for regional-scale mapping and monitoring initiatives.

2 Methods

2.1 Study Site

The site selected for this study is the Tunbridge Township Lagoon Reserve, located in the Tasmanian Midlands Region approximately 100 km from the Capital of Hobart. The 20 ha reserve is the only designated protected area in the State for many endangered grassland communities, and has therefore been the target of detailed vegetation studies in the past (Zacharek *et al.*, 1997). The site contains a shallow saltwater lagoon, surrounded by native grassland communities. The eastern third of the site contains a small hill intergrading into an open grassy woodland on neighbouring properties. This site has been selected as it exhibits excellent examples of major lowland native grassland community types within a small area, and is free from factors that may confound confusion between vegetation classes, such as grazing and fertilisation.

2.2 Data Collection

In November 2015, a field campaign was carried out to collect the data used in this study. A random sample of 55 sites was created, and each site surveyed. At each site, a 5 m x 5 m star transect was created by running a measuring tape north to south and then east to west with the random coordinate of the site situated at the transect centre. Spectral signatures were collected using an ASD Handheld-2 spectroradiometer (ASD Incorporated, 2010) which collects signatures between 375 nm and 1,075 nm in 1 nm increments. Vegetation types were recorded and photographed at the transect centre, and at 2.5 m and 5 m in each compass direction along the transect for a total of 9 observation points. At each observation point, three spectral signatures were collected at nadir and off-nadir angles in order to record all variation in species composition and structure. Collection of narrowband spectral measurements in such a way has been shown to reduce uncertainty in measurements of biological and geographical parameters derived from image spectra (Gianelle and Guastella, 2016). The integration time for the spectroradiometer was optimised at each new site, and a dark current subtraction and a white reference observation (of a Spectralon panel) was performed at each observation point along each transect. This was done in order to i) reduce drift and noise of the spectroradiometer, ii) reduce potential variation between sites as a result of varying solar angles throughout the day, and iii) reduce potential issues arising from cirrus cloud cover common to the area. In addition, each reading from the sensor was averaged over 20 observations in order to reduce noise. Figure 3.1 shows the location of each transect throughout the site. Table 3.1 shows the number of field samples collected for each class.

Table 3.1: Number of training and validation points used per class. Total point count is given in the final column. The C_3 class count is the sum of the *Danthonia* and *Wilsonia* counts.

Class	Training Points	Validation Points	Total
Themeda	316	163	479
Danthonia	213	109	322
Wilsonia	197	102	299
C_3	410	211	621
Saltpan	193	100	293

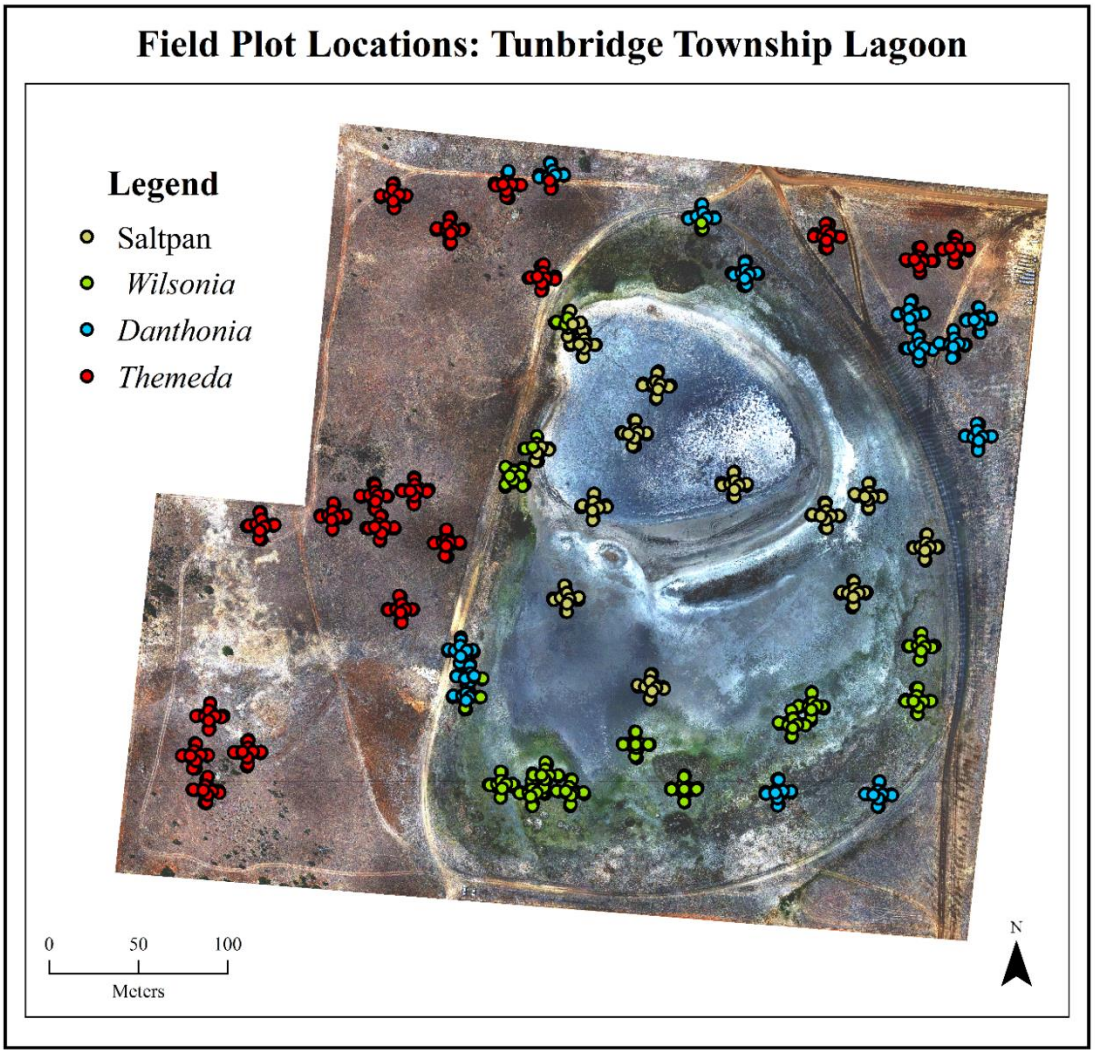

Figure 3.1: Location of field plots showing transect locations and observation points. The associated vegetation class are for each observation is shown as per the legend

2.3 Class Descriptions

For this study, four classes were defined: 1) saltpan, 2) *Wilsonia rotundifolia/ Selliera radicans,* 3) *Danthonia sp.*, and 4) *Themeda triandra.* These classes are loosely based on the floristic communities identified by (Zacharek *et al.*, 1997), who proposed two additional communities;

Calocephalus lacteus open grassland, and *Lolium perenne* grassland which were excluded due to insufficient data points and limited observable extent. The saltpan class covers areas of exposed soil which have no vegetative cover, and includes both the lagoon itself and the surrounding mud flats. The *Wilsonia rotundifolia/Selliera radicans* class covers areas of saltpan where the dominant cover consists of either of these two species, with no grass species being present. This class is similar to the *Puccinella stricta* grassland class proposed by (Zacharek *et al.*, 1997). The *Danthonia sp.* class is quite broad, covering any area that contains *Danthonia* sp. or *Poa* sp. (such as *P. rodwayi* or the larger *P. labillardieriei*) regardless of the inter-tussock species. The definition of this class is similar to that of the grassland complex class found in the TASVEG community listings, and represents one of the three major lowland native grassland community types. This class covers areas from the edge of the lagoon to the foot of the hill, which composes the eastern third of the site. There is significant variation in secondary inter-tussock species throughout the class, and intergrading with both the *Wilsonia* class and the *Themeda* class is in some places extensive. The main distinguishing factor for this class is that the dominant grass species are all cool season C_3 grasses, exhibiting a typical winter-spring growth period, and a period of senescence over the summer months. The final class, *Themeda triandra*, represents a typical lowland native grassland community, and covers areas located in the eastern half of the site. *Themeda triandra* is dominant in these areas, with some dispersed trees and shrubs, primarily *Acacia dealbata* and *Bursaria spinosa.* (Zacharek *et al.*, 1997) proposed that the remnant *Themeda* communities found on the site originally formed an open grassy woodland, due to occurrence of *Eucalyptus ovata* specimens on neighbouring properties. The distinguishing feature of this class is that the dominant species is always *Themeda triandra,* which is a warm season C_4 grass, typically growing in the warmer summer months and entering senescence in late autumn to winter.

2.4 Datasets

In order to determine the spectral resolutions required for accurate classification of different lowland native grassland communities, a number of tests were run. A total of four different datasets were derived from the original HSR field data. The first dataset used the full set of HSR bands in their original 1 nm increments, resulting in a total of 501 input bands. The second dataset used a variable reduction process to identify redundant bands within the original 501 band dataset. One of the major issues associated with high dimensionality spectral datasets is the occurrence of multiple co-linearity between bands (Darvishzadeh *et al.*, 2008; Adjorlolo *et al.*, 2012). Due to the high number of input bands, the number of samples required to establish a statistically meaningful result from subsequent image analysis can become exceedingly high (Clevers *et al.*, 2007; Adjorlolo *et al.*, 2012). This problem is commonly referred to as the Hughes phenomenon, or the 'curse of dimensionality'. In order to reduce issues associated with the Hughes phenomenon and multiple co-linearity between bands, variable reduction techniques are commonly used. Irisarri *et al.* (2009) found that the most common method used to mitigate such issues within high spectral resolution approaches was to first transform the data, and then run a feature selection protocol in order to identify and remove data redundancies. Clevers *et al.* (2007), however, warned against the transformation of data before feature selection, as when data has been transformed, the ability to interpret outputs within a physical and environmental context is lost. In this case, it was decided to omit the data

transformation stage, as the ability to relate class reflectance properties to plant biophysical characteristics was highly desirable. Therefore, only a feature selection protocol was applied.

Many authors have proposed methods for reducing data redundancy through feature selection processes (Schmidt and Skidmore, 2003; Clevers *et al.*, 2007; Pal and Foody, 2010; Adjorlolo *et al.*, 2012). In order to reduce redundancy within highly correlated spectral bands, the dataset was run through the GeneSrF package in R (R Core Team, 2015). GeneSrF is a variable reduction protocol that utilises random forests as a means of identifying non-redundant variables with strong predictive capabilities (Diaz-Uriarte, 2007). A key strength of random forest models is that they can provide estimates of variable importance for the resulting model derived from the training data (Breiman, 2001), however, the variables identified as having high importance values are often strongly correlated (Archer and Kimes, 2008), which must be considered in the final interpretation of results. The GeneSrF approach works by first iteratively excluding a predetermined percentage of variables used in the previous iteration with the lowest importance scores (typically 20%). This process is repeated until all trees are fitted to the dataset (Diaz-Uriarte, 2007). The trees are then examined, and the tree with the smallest number of included variables that still has an out-of-bag error estimate lower than a user-determined threshold between 0 and 1 is identified. The variables used in this tree are then extracted and used in the reduced model (Diaz-Uriarte, 2007). The protocol was run independently over the original HSR data for both the three-class and four-class versions of the dataset. A total of 106 spectral bands were identified as belonging to the optimal model for the three-class configuration, and 86 spectral bands for the four-class alternative.

In order to determine if high spectral resolution input data is required to differentiate between the lowland native grassland communities, or whether broadband spectral approaches are sufficient, the original narrowband HSR dataset was resampled to match the spectral resolution of the Landsat 8 and WorldView-2 sensor platforms, using a spectral convolution method analogous to the spectral resampling workflow in ENVI 5.2 (Exelis Visual Solutions, 2013) (Cundill *et al.*, 2015; Roth *et al.*, 2015). Due to the limited spectral range covered by the ASD Handheld 2 spectroradiometer, not all bands were able to be simulated for all sensors. For sensor bands in which there was no recorded field data, the bands were excluded from analysis. For Landsat 8, only the first 5 bands (430-880 nm) could be emulated, while for WorldView-2 the first 7 bands could be emulated (400-895 nm). The convoluted Landsat 8 and WorldView-2 spectra were then assigned class labels using both the three and four-class configurations. Figure 3.2 shows the mean spectral signature for each class in each dataset. The three-class sand four-class reduced high spectral resolution dataset values are provided on separate plots as the two datasets have different optimal band selections.

2.5 Classification

Classification was undertaken using a random forest (RF) approach similar to that used in Chapter two. Due to the low number of samples available for each class, and high spectral variability within classes, a *k*-fold cross-validation approach was employed. *K*-fold cross-validation is a method used to ensure that the process of subdividing reference datasets into

training and validation subsets does not introduce bias into the classification result. As the RF classification approach sets aside 33% of the training data used to build each tree for cross-validation, and as a result the number of training points used to generate the classification model is significantly reduced from the original number. This can create issues in cases where the initial number of input points is low, as it means that the model cannot be drawn based on the full variability and range of the collected dataset. In the *k*-fold validation approach, classification is undertaken multiple (*k*) times based on different random splits of the reference dataset. For each randomly generated subset, a single RF model is created using the points identified for training, and then the resulting classification evaluated against the reciprocal points. This process is repeated multiple times, and average training and final classification accuracies determined for each class. By repeating the classification and validation approaches multiple times using various random subsets of the data, RF models are trained across the entire range of the dataset, and therefore it can be determined if there are significant variations in accuracy based on the configuration of training and validation points. For this study, a series of 30 random subsets were produced from the original dataset using the scikit-learn module in Python using a 66% training to 33% validation distribution (Pedregosa *et al.*, 2011).

Each of the 30 subsets was used to train, classify and validate a RF model based on each of the datasets using both a three and four-class configuration. This resulted in a total of 8 classification results, drawn across 30 repeat classifications. For each of the results, the number of variables to try was set equal to the square root of the number of input bands (Breiman, 2001; Naidoo *et al.*, 2012). For each of the output results, training and validation accuracies were averaged across the 30 repetitions, and reported for each class. Variable importance measures were also derived from the RF training models, and averaged for each result.

One of the most important goals of this study was to determine whether there are significant differences in classification accuracy for the analysed lowland native grassland communities based on the input spectral resolution of the dataset. A secondary goal was to determine whether merging the C_3 vegetation into a single class improved classification accuracy, as similar types of vegetation exhibited similar confusion rates in the results obtained in Chapter 2. Therefore, to determine whether such effects exist, a series of analysis of variance (ANOVA) tests were undertaken. Differences in mean classification accuracy were tested within classes across the range of datasets, and additionally between class configurations within each dataset.

3 Results

3.1 RF Training Accuracies

Average training accuracies acquired for all three-class results are summarised in table 3.2. Percentages are obtained by averaging per-class results across all 30 trials. There is a high degree of similarity in class results across the different datasets. Mean accuracies for the merged C_3 class in particular are extremely close. Standard deviation values are small for all results, at approximately ±1%. All class accuracies are high, with the lowest accuracy being 85.2% for the C_3 class in the full HSR, reduced HSR and Landsat 8 results, and for the *Themeda* class in the Landsat 8 result.

Table 3.2: Average RF training accuracy for all three-class results. Accuracies are given as a percentage obtained by averaging the results of all 30 RF training models. Standard deviations are given as a percentage value above or below the mean.

	Full HSR	Reduced HSR	Landsat 8	WorldView-2
Saltpan	93.6 ± 0.9	93.3 ± 1.1	93.3 ± 1.0	93.9 ± 1.0
C_3	85.2 ± 1.1	85.2 ± 1.1	85.2 ± 1.0	85.1 ± 1.3
Themeda	86.3 ± 0.9	85.4 ± 1.2	85.2 ± 1.4	85.7 ± 1.1
Overall	87.3 ± 0.6	86.9 ± 0.7	86.9 ± 0.8	87.1 ± 0.8

Table 3.3 reports the average RF accuracies obtained from the range of four-class trials. Accuracies for these results are more variable than for the three-class results. The full HSR model obtains the highest accuracies for all vegetation classes, and additionally has the lowest standard deviation. The WorldView-2 result has the highest classification accuracy for the saltpan class. Standard deviations have increased over the three-class results, most noticeably for the two classes composing the C_3 class; *Wilsonia* and *Danthonia*. The *Danthonia* class has poor results in all datasets, while other classes have comparable or only slightly decreased accuracies from the three-class results.

Table 3.3: Average RF training accuracy and standard deviations for all four-class results, as averaged across the 30 RF models. Accuracies are presented as mean percentages, while standard deviation is presented as a percentage range above or below the mean.

	Full HSR	Reduced HSR	Landsat 8	WorldView-2
Saltpan	93.8 ± 1.1	92.6 ± 1.3	93.8 ± 0.9	94.1 ± 0.9
Wilsonia	80.9 ± 2.1	77.7 ± 2.2	78.4 ± 2.3	79.1 ± 1.9
Danthonia	59.1 ±2.4	56.1 ± 2.6	56.9 ± 2.4	56.9 ± 2.7
Themeda	88.2± 1.1	88.8 ± 1.5	87.0 ± 1.2	87.9 ± 1.0
Overall	81.1 ± 0.8	79.6 ± 0.8	79.6 ± 0.8	80.1 ± 0.9

3.2 RF Variable importance measures

Figure 3.3 shows the mean variable importance measures for each set of three-class trials. The full and reduced HSR values, shown in the top row of the figure, show patterns of selection for variables in the 400 nm, 550 nm, and 675 nm portions of the spectrum. The reduced HSR model, in contrast to the full HSR model, exhibits increased importance for the bands retained in the 770 nm and 890 nm portions of the spectrum. The two broadband results show contrasting selections again, with the Landsat 8 model showing high importance in band 4 (640-670 nm), for all vegetation classes, before a decrease in importance for band 5 (850-880 nm). The WorldView-2 model shows peak importance in band 5 (630-690 nm), and bands 1 (400-450 nm) and 7 (770-895 nm) also being identified as important.

Mean variable importance scores for each of the four-class results are summarised in figure 3.3. High importance variables in for all results are very similar to those selected in the three-class trials. Key spectral regions are again associated with similar wavelengths, such as at 400

nm, 550 nm, and 675 nm in the HSR results. The reduced model again shows increased importance in the near infrared region (800 nm).

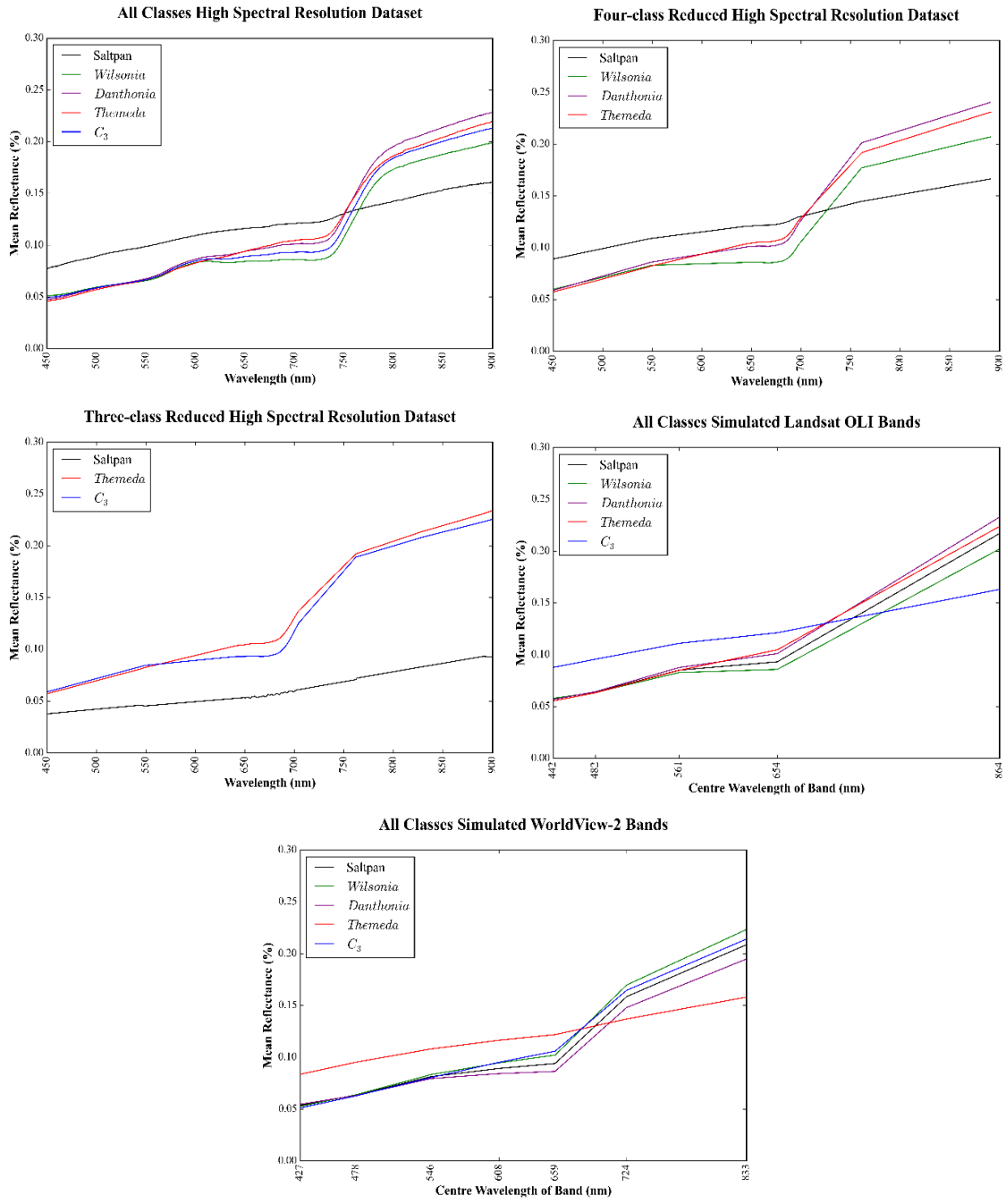

Figure 3.2: Mean spectral signatures for all classes in each dataset. Landsat-OLI and WorldView-2 mean spectra are plotted against the centre wavelength of each respective band.

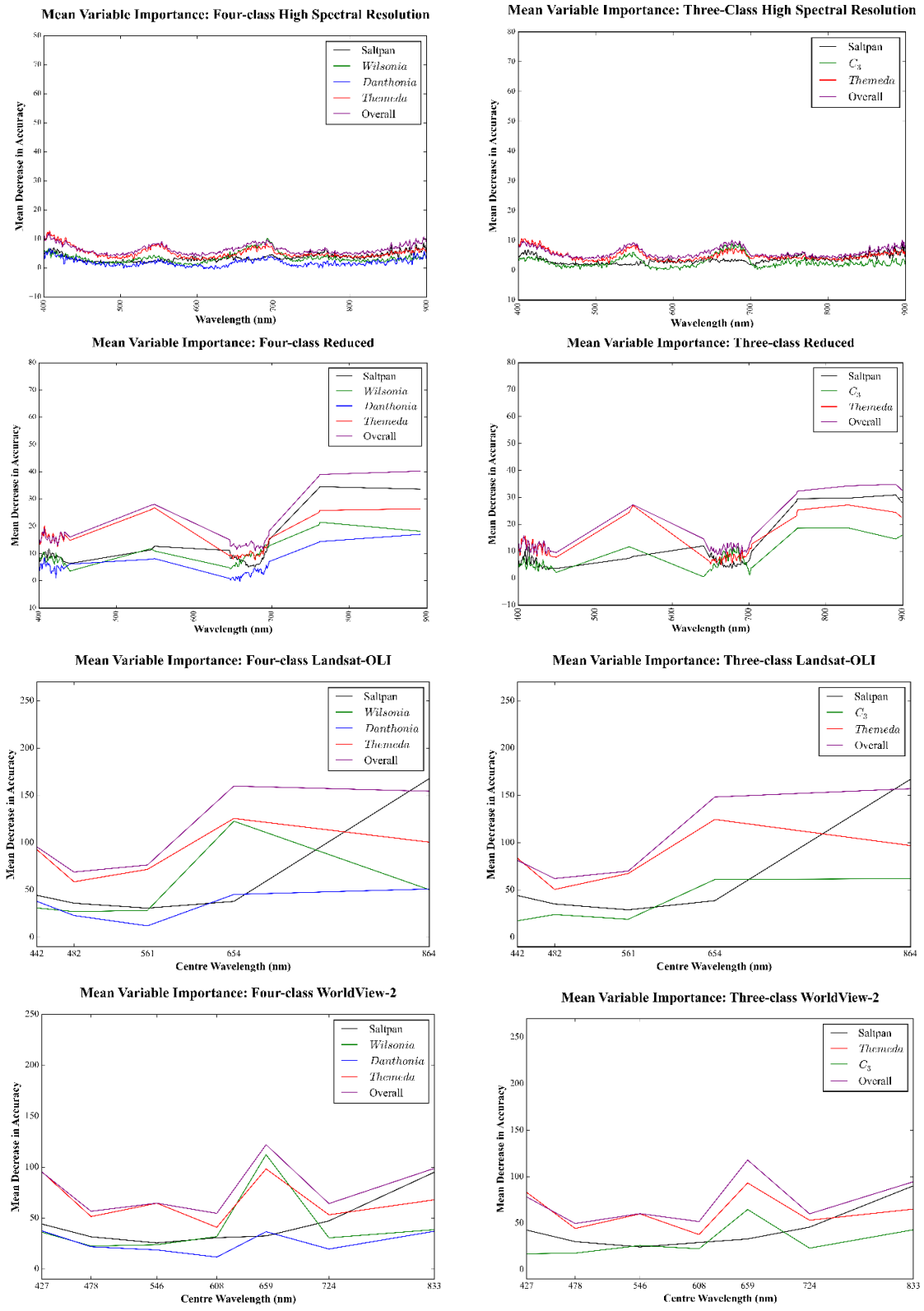

Figure 3.3: Mean RF variable importance for all class configurations and datasets. Importance values are reported for each band used in the set of trial, and averaged across the set

3.3 Final classification accuracies

Table 3.4 shows the final mean classification accuracies and confusion matrix for the full HSR three-class results. The overall accuracy for the dataset was 87.4% ±1.2. All classes show high mean accuracies, and confusion is limited between classes.

Table 3.4: Mean final confusion matrix for full HSR three-class tests

	Saltpan	**C_3**	***Themeda***	**User's Accuracy**
Saltpan	92.7	3.0	3.3	93.6 ± 2.7
C_3	5.9	193.5	26.6	85.6 ± 1.8
Themeda	2.1	20.6	140.2	86.0 ± 2.1

Table 3.5 shows the final mean classification accuracies and confusion matrix for the reduced HSR three-class results. The overall mean accuracy achieved was 86.8% ± 1.1. Class accuracies are similar to those obtained from the full HSR three-class model, and the degree of confusion between classes is also consistent between the two results.

Table 3.5: Mean final confusion matrix and classification accuracies for the reduced HSR three-class tests

	Saltpan	**C_3**	***Themeda***	**User's Accuracy**
Saltpan	91.6	2.7	4.6	92.6 ± 2.5
C_3	5.8	193.3	27.0	85.5 ± 1.8
Themeda	1.8	22.5	138.8	85.1 ± 2.9

Mean classification accuracies for the Landsat 8 three-class results are show in table 3.6. The overall mean accuracy was equal to 86.8% ± 1.5. Accuracies are again very similar to those obtained in the previous three-class tests.

Table 3.6: Mean final confusion for the Landsat 8- OLI three-class tests

	Saltpan	**C_3**	***Themeda***	**User's Accuracy**
Saltpan	92.6	2.5	3.9	93.5 ± 2.7
C_3	6.0	192.9	27.1	85.3 ± 2.2
Themeda	1.6	23.1	138.3	84.8 ± 2.6

Table 3.7 shows the mean final classification accuracies for the WorldView-2 tests. Mean overall accuracy was equal to 87.1% ± 1.4. Class-specific accuracies are within similar ranges to those obtained from the previous three-class results, as are confusion rates and standard deviations.

Table 3.7: Mean final confusion for the WorldView-2 three-class tests

	Saltpan	**C_3**	***Themeda***	**User's Accuracy**
Saltpan	92.5	2.8	3.7	93.4 ± 2.8
C_3	5.6	193.5	27.0	85.6 ± 1.9
Themeda	1.9	22.0	139.1	85.4 ± 2.5

Classification results for the four-class reduced hyperspectral tests are reported in table 3.8. Overall accuracy was equal to 80.1% ± 1.7. Accuracies of all classes except for *Themeda* have decreased slightly from the full HSR four-class model. The majority of confusion is related to the *Danthonia* class, where misclassification occurs as a result of confusion with both the *Themeda* and *Wilsonia* classes

Table 3.8: Final mean classification accuracies for the four-class HSR tests.

	Saltpan	***Wilsonia***	***Danthonia***	***Themeda***	**User's Accuracy**
Saltpan	92.9	1.8	0.6	3.7	93.8 ± 2.6
Wilsonia	4.9	96.0	13.2	2.9	82.0 ± 4.2
Danthonia	1.4	15.8	65.2	26.6	59.8 ± 5.1
Themeda	2.1	1.3	16.5	143.1	87.8 ± 2.6

Table 3.9 gives the final mean classification accuracies and confusion matrix for the four-class reduced HSR tests. There is an increase in mean accuracy for the *Themeda* class to 89.6% ± 2.4, however, all other results are similar to those obtained previously

Table 3.9: Final mean RF classification averages for the reduced HSR four-class tests

	Saltpan	***Wilsonia***	***Danthonia***	***Themeda***	**User's Accuracy**
Saltpan	92.8	1.2	0.9	4.1	93.8 ± 2.7
Wilsonia	5.5	93.6	15.7	2.2	80.0 ± 4.4
Danthonia	1.4	17.5	61.8	28.2	56.7 ± 4.0
Themeda	1.5	1.7	18.7	141.0	86.5 ± 2.3

Table 3.10 shows the final mean classification accuracies and confusion matrix for the four-class Landsat 8 tests, showing similar degrees of accuracy to previous four-class tests. Finally, mean classification accuracies for the WorldView-2 four-class trials are shown in table 3.11. Obtained accuracies are highly similar to those obtained from previous four-class tests. The total overall accuracy was equal to 80.2% ± 1.9.

Table 3.10: Final mean classification accuracies for the Landsat-OLI four-class tests

	Saltpan	***Wilsonia***	***Danthonia***	***Themeda***	**User's Accuracy**
Saltpan	92.6	1.6	0.7	4.0	93.6 ± 3.0
Wilsonia	5.0	94.1	14.8	3.0	80.4 ± 4.6
Danthonia	1.4	17.1	61.9	28.5	56.8 ± 4.1
Themeda	1.8	1.3	17.3	142.6	87.5 ±2.4

Table 3.11: Final mean classification accuracies for the WorldView-2 four-class tests

	Saltpan	***Wilsonia***	***Danthonia***	***Themeda***	**User's Accuracy**
Saltpan	92.9	1.8	0.6	3.7	93.8 ± 2.6
Wilsonia	4.9	96.0	13.2	2.9	82.0 ± 4.2
Danthonia	1.4	15.8	65.2	26.6	59.8 ± 5.1
Themeda	2.1	1.3	16.5	143.1	87.8 ± 2.6

3.4 ANOVA results

The ANOVA test undertaken to determine whether class-specific classification accuracies vary based on the dataset indicated that for the three-class tests, there was no significant difference in classification accuracy for any class. This means that there is no significant reduction or improvement in classification performance between datasets based on spectral resolution. For the four-class tests however, it was found that the *Themeda*, *Danthonia*, and overall classification accuracies did have significant variations in accuracies based on the input dataset, at $p \leq 0.05$. The results of Tukeys' post-hoc comparisons indicated that for the *Themeda* class, the reduced HSR model result had statistically significant higher mean classification accuracy than any of the results. For the *Danthonia* class, and the overall classification accuracy, it was found that the Landsat 8 and WorldView-2 accuracies were significantly lower than the accuracies obtained in the full HSR model. There were, however, no other significant differences detected between other datasets.

The second set of significance tests undertaken was used to determine whether differences between class means could be detected within the results of each dataset. The results indicated that the Landsat 8, Worldview-2 and full HSR datasets exhibited significantly higher classification accuracy for the combined C_3 class over both the *Danthonia* and *Wilsonia* classes used in the four-class tests. Additionally, the accuracy of the *Danthonia* class was significantly poorer than the accuracy of the *Wilsonia* class in all of the above datasets. For the reduced HSR tests, the *Themeda* class from the four-class model has significantly higher accuracy than the three-class equivalent. Overall accuracy across all classes, however, was determined to be significantly higher than the

overall four-class accuracy. Similar to the Landsat and WorldView-2 results, the C_3 class in the reduced HSR results was shown to have better classification accuracy than both the *Danthonia* and *Wilsonia* classes, and *Wilsonia* was shown to have higher mean accuracy than *Danthonia.*

4 Discussion

4.1 Training and classification accuracies

Mean training and classification results obtained for all combinations of classes and datasets show high degrees of similarity. Class-specific accuracies typically vary from the training accuracy achieved within a dataset by approximately 1%. The observed similarities in accuracy indicate that the sampling protocol is robust, and that full class variation has been accounted for in both the training and validation stages of classification. The low standard deviations observed in both training and classification accuracies again indicate consistency in the results, and show that the distribution of field plots and subsequent division of data into training and validation subsets is unbiased.

Final classification accuracies achieved for the range of three-class tests are good, with classes exhibiting consistent behaviour across the four datasets. Accuracies are very similar within classes for the various datasets, with standard deviations also showing highly consistent values. Confusion rates are similarly consistent across the range of tests, with average pixel counts between classes differing by less than 1. The primary source of misclassification is confusion between the C_3 and *Themeda* classes. The rate of misclassification is slightly higher for *Themeda* points being misclassified as C_3 than for C_3 being misclassified as *Themeda*. Confusion within the saltpan class is negligible, with very few points in either of the other two classes being wrongly attributed to this class.

The results for the four-class tests also exhibit consistent levels of accuracy across the range of datasets, although there is slightly more variation observable in the three vegetation classes, although such differences in accuracy are only in the order of ~3%. Confusion between classes is similarly for both the three and four-class results. Classification accuracy for the *Danthonia* class is significantly poorer than the results obtained for the other classes, ranging from 57% for the Landsat 8 tests to 59.8% for the full HSR tests. The cause of these poor results appears to be due to consistent confusion with the *Themeda* class, observable across all four datasets. For each of the four datasets, approximately 25% of all *Danthonia* validation and training points are consistently identified as *Themeda* across the 30 subsets. Confusion with the *Wilsonia* class is also a significant contributor to the poor accuracy of the *Danthonia* class, with approximately 12% of points being misclassified in this manner. The higher rate of confusion between *Danthonia* and *Themeda* is likely due to similarities in canopy structure between the two classes, as both exhibit a typical erectophile morphology that is not present in the *Wilsonia* class. The primary physiological similarity between the *Wilsonia* and *Danthonia* classes is the shared photosynthetic pathway, which will be primarily expressed spectrally as higher levels of greenness than observed

in the C_4 *Themeda* class due to differences in phenological staging. Confusion for the *Wilsonia* class is almost exclusively with the *Danthonia* class, with a very few points being misidentified as *Themeda.* Confusion within the *Themeda* class is consistently with the *Danthonia* class, again likely as a result of similar canopy structure.

4.1 Variable importance measures

Figure 3.3 shows the variable importance measures for the three-class tests. Prioritisation of spectral regions is similar between the four datasets. Key wavelengths identified are at 400 nm, 550 nm, 675 nm, and 900 nm. For the two HSR tests, the reduced model has significantly higher importance levels than the full HSR model with wavelengths 675 nm, 780 nm and 890 nm being identified as highly important. The full HSR model shows lower importance values for longer wavelengths than those observed in the other three-class models. The reduced model has high numbers of variables selected in the regions surrounding 400 nm and 675 nm, indicating the presence of uncorrelated information contained within these bands. The smaller number of variables selected by the reduction process in the remaining key regions are, however, more important overall to the classification models. In the broadband Landsat 8 and WorldView-2 tests, the most importance bands cover similar spectral regions as those identified in the HSR results. For the Landsat 8 result, the most important band is band 4 (640-670 nm), followed by band 5 (850-880 nm). For the WorldView-2 result, the most important band was band 5 (630-690 nm), followed by band 7 (770-895 nm) and band 1 (400-450 nm). The Landsat 8 result is the only model that does not indicate localised high importance levels in the shorter wavelengths, as the sensor does not cover the 400-450 nm region, while all other sensor configurations do.

Within the three-class results, class-specific variable importance measures indicate the presence of key regions of separability. The *Themeda* class exhibits much higher importance compared to the C_3 class in all regions. The area in which this difference is most evident is in the shortwave 400-450 nm region, in which the C_3 class exhibits only a slight increase in importance, while the *Themeda* class shows a clear increase. In the reduced HSR model, the *Themeda* class has clearly higher importance levels at 550 nm, and 680 nm. These spectral wavelengths are strongly associated with plant pigment levels, most notably carotenoids in the 400-450 nm region (Zur *et al.*, 2000), and anthocyanin at approximately 550 nm (Sims and Gamon, 2002). Increased reflectance in regions associated with such pigments is a good indicator of vegetation senescence (Gitelson and Merzlyak, 1994). Given the varying phenological staging of the two classes as a result of their different photosynthetic pathways, these associations are not unexpected. The reduced HSR model, Landsat 8 and WorldView-2 models all indicate strong variable importance at 890 nm for all classes, including the soil class. This region is known to be associated with water content (Tucker, 1977), which again will differ between vegetation types as a result of phenological differences.

The variable importance results obtained for the four-class tests identify the same spectral regions as being of key importance as the three-class results. Band selections are almost identical for the reduced three and four-class models, although fewer bands around 680 nm were selected for the four-class results. In all of the datasets, the *Themeda* class again shows high importance values at 400-450 nm, as well as at 550 nm in the two HSR trials. The *Wilsonia* class shows high importance in the broadband Landsat 8 and WorldView-2 trials in the same bands as identified as important for the C_3 class, however the degree of importance is dramatically higher than in the three-class equivalent tests. The *Danthonia* class shows low importance across the range of spectral regions, although localised peaks can be detected at 775 nm and 890 nm within the reduced HSR result, in band 1 (400-450 nm) and band 5 (630-690 nm) for WorldView-2 result, and bands 4 (640-670 nm) and band 5 (850-880 nm) in Landsat 8. The selection of the same spectral regions as identified in the three-class tests indicates that the same plant biophysical properties as discussed previously are likely to be the key drivers of class separability in both the four and three-class results.

4.2 ANOVA and Tukey's post-hoc comparison results

As each dataset was classified multiple times as a result of employing a *k*-fold cross-validation approach, statistical measures indicating the significance of differences in classification accuracy observed between datasets and classes can be produced. The production of such measures is important to this study, as it allows for accurate determination of the spectral resolutions required to accurately classify lowland native grassland communities, as well as optimal class configurations for improved classification outcomes.

For the three-class tests, the ANOVA analysis indicated that there were no significant differences in classification accuracies between datasets. This means that no class has significantly improved or reduced mean accuracy when classification is performed using different datasets. Therefore, it can be concluded that the two vegetation classes, C_3 and *Themeda*, can be differentiated with the same degree of accuracy using both HSR and broadband datasets. The results obtained indicate that HSR datasets may provide little improvement in lowland native grassland community differentiation over broadband multispectral approaches under these conditions. For the four-class results however, the ANOVA analysis identified several key differences in class accuracy across the range of datasets. For the *Themeda* class, a significant improvement in accuracy within the reduced HSR model was detected. Additionally, the *Danthonia* class was identified as having significantly poorer accuracies in the Landsat 8 and WorldView-2 results compared to the reduced HSR model, however no difference in accuracy was detected between the full and reduced HSR models. Overall classification accuracy for all classes followed a similar pattern, with poorer performance in the Landsat 8 and WorldView-2 results compared to the reduced HSR output. This finding indicates that when the C_3 class is split into *Wilsonia* and *Danthonia* classes, there is a distinct need for HSR datasets in order to ensure classification accuracy.

When differences in class-specific accuracy were analysed within each dataset, the results based on the Landsat 8, WorldView-2 and full HSR models were similar. No significant variation in classification accuracy between the three and four-class versions of the *Themeda* and saltpan classes respectively were identified, meaning that differences in final accuracy were non-significant within these datasets based on the class configuration. The results for these three datasets indicate that the combined C_3 class has significantly improved classification accuracy over the *Wilsonia* and *Danthonia* classes. The results for the analysis within the reduced HSR model identified that there was a significant increase in classification accuracy for the *Themeda* class based on the four-class class configuration over the three-class alternative. Additionally, the results indicated that again, the C_3 class has significantly higher classification accuracy than both the separated *Wilsonia* and *Danthonia* classes within this dataset, and that the overall accuracy for the three-class result is significantly higher than the overall four-class accuracy.

Overall, the results of the ANOVA indicate that for the majority of classes, there is no improvement in classification accuracy based on the dataset, however, the application of a simplified three-class model does result in significantly higher general classification accuracy. The *Themeda* class, in contrast to these generalised findings, reaches optimal performance based on a full four-class model employed on a reduced HSR dataset. Even though the generalised results indicate that a broadband three-class model is likely to be the best candidate for generalised community differentiation, the inability of such models to accurately discriminate between the *Danthonia* and *Themeda* classes is of concern. As the *Themeda* class is a key class in this study, it is concluded that optimal model selection should prioritise good performance within this class over the performance of others.

4.3 Spatial considerations

As the field sampling design used in this study is randomised and based on clustered observations along transects, the number of observations collected for each class is not equal. In the four-class model, a total of 214 *Danthonia* points and a total of 228 *Wilsonia* points were identified. When combined in the three-class model, this results in a total of 442 C_3 training points. As RF models automatically set aside 33% of all input training points used to grow a tree for cross-validation (Breiman, 2001), this reduces the number of potential training points for both classes significantly, to 153 points for the *Wilsonia* class, and 143 points for the *Danthonia* class. In comparison, when 33% of all C_3 training points are withheld, the model is still created using a total of 296 points for the class. The inclusion of fewer points in the training stage of RF classification can result in significantly poorer final outcomes, as the random subsetting process may derive a non-representative sample from the larger input training dataset to grow the tree from. This may result in trees being unable to accurately classify the data in the final stage of analysis. Therefore, the C_3 class has an inherent advantage over the two split classes, potentially resulting in artificially inflated classification accuracies. The occurrence of this issue could be addressed in future studies through the employment of a stratified field sampling protocol, or by randomly selecting a subset

of the C_3 class with a similar number of input points to the *Wilsonia* and *Danthonia* classes in order to ensure equal sample sizes. In order to determine whether the effect of unequal sample sizes has adversely affected the results for the four-class trials, further analysis needs to be undertaken.

The analysis in this study is focussed solely on determining the spectral resolution and class configurations required for accurate community classification, and as a result does not consider spatial resolution and its effects on classification results. As the ASD handheld-2 spectral radiometer has a 25° field of view, based on a one-meter sampling height, spectra are collected across a circular area with a diameter of 46 cm. Care was additionally taken to ensure that spectra covered patches only consisting of a single class. Even though the results obtained in this study, when taken in conjunction with the results of Chapter 2, indicate that broadband approaches to lowland native grassland differentiation are feasible, the comparatively coarse spatial resolution of the Landsat 8 and WorldView-2 sensors may have a significant effect on final accuracy. The mixture of multiple thematic classes within a single pixel will result in significantly higher deviation in class spectral signatures, and additional increases in class generalisation. The results for the simulated broadband sensor tests in this chapter assume spectral signatures to be representative of only a single class. The results obtained in Chapter 2 indicate that classes can be accurately identified using such broadband approaches, however, the classification accuracies that were obtained are much lower than the results presented in this chapter. The deviation in accuracy observed for classes between the two studies undertaken thus far is likely to be the result of varying spatial resolutions and the sampling methods used to generate the training and validation datasets. The use of field-based methods for point collection in this case is likely to be more accurate than the method used in the previous chapter. Each sample collected in the field was individually visited multiple times, and the land cover class confirmed. This contrasts to the approach used in the previous chapter, in which polygon extents were generated based on extrapolation of field observations. The use of a more spatially precise sampling regime is likely to have contributed to the superior accuracies achieved in this study compared to the results reported in Chapter 2.

5 Conclusions

The results of this study provide several key findings that can be used to improve future lowland native grassland mapping approaches. Firstly, the results from the series of three-class tests indicate that for generalised class configurations, broadband spectral resolutions are capable of providing high classification accuracies for lowland grassland communities. Final accuracies for the *Themeda* class in the broadband results ranged between 84.8% for Landsat ETM+ and 85.4% for WorldView-2. The results of the ANOVA indicate that there is no significant improvement in classification accuracy when analysis is undertaken using HSR datasets for the three-class configuration. The second major finding is that the separation of *Danthonia* and *Themeda* grasslands is not feasible based solely on spectral properties. All of the spectral resolutions trialled in this study failed to provide accurate differentiation between these two classes, indicating that

future classification results must consider alternate variable sources to provide good results. The inclusion of variables related to habitats, nutrient status, phenology or structure may provide some improvement on the results obtained here. The improvement in classification accuracy for the merged C_3 class to 85% mean accuracy from 80% mean accuracy for the *Wilsonia* class and 57% mean accuracy for the *Danthonia* class was found to be statistically significant on $p < 0.05$. The third key finding of the study is that *Themeda triandra* communities have significantly higher classification accuracy in is the HSR results, in which the mean classification accuracy was 89%, compared to broadband results in which average accuracy was 86%. Additionally, the *Themeda* class benefits from the separation of the *Wilsonia* and *Danthonia* classes, with a statistically significant improvement in classification accuracy from 84% in the three-class results, to 87% in the four-class. This result indicates that the use of HSR datasets is valuable in this case, as even though the accuracy of other classes is not improved over broadband approaches, there is no indication that using HSR datasets results in poorer classification outcomes over broadband datasets. The final key finding of this study relates to variable importance measures and their relationships to plant biophysical properties. The importance measures obtained across both sets of classes and all datasets indicate key regions of separability being related to pigment levels and water content. These associations provide valuable insight into the communities, and can aid in the selection of appropriate spectral ranges for future studies, and in the selection of optimal data collection times.

Overall, the results of this study indicate that classification of lowland native grassland communities using HSR datasets is possible. The key findings of the work provide valuable insight and information that can be used to improve future mapping approaches. The results additionally corroborate the findings of previous chapters, indicating that there are significant issues associated with classification of C_3 grass species, and that the incorporation of non-spectral variables is likely to be important for ensuring accurate results.

6 Thesis context

The results of this chapter identify the spectral resolutions required to differentiate lowland native grassland communities. The work in this chapter has determined the key spectral regions in which class separability is optimised. These findings address issues raised in Chapter 2, and provide tangible solutions. Datasets meeting the required spectral and spatial resolutions identified across Chapters 2 and 3 can now be employed in order to provide potentially improved classification results.

Chapter 4

Prediction of Optimal Segmentation Scale on a Per-Class Basis Using Combined Thematic and Spatial Metrics

Abstract

The selection of the optimal segmentation scale for object-based image analysis techniques has proven to be a difficult problem to overcome. Many authors have provided insight into the issue, and proposed methods for determining such an ‘optimal’ scale based on spatial difference metrics applied across a scene. There has not, however, been any method capable of determining class specific optimal segmentation scale, or any that successfully incorporates thematic elements into the determination of the optimal scale factor. In this chapter, a novel approach is outlined, that is capable of predicting the optimal image segmentation scale through the use of combined spatial and thematic performance metrics. The approach uses a set of indices designed to measure thematic similarity between individual segments and target classes, in conjunction with the degree of undersegmentation. The metrics were implemented and tested in two case studies: an urban environment with clearly defined class boundaries, and a lowland native grassland community patch located in the Tasmanian Midlands. These different case studies were chosen in order to test the performance of the metrics in environments with both clear object boundaries and in a more natural environment with less discrete boundaries. The determination of the optimal segmentation scales was previously identified as a pertinent issue in the identification of high conservation priority lowland native grassland patches, and therefore the development of this approach is designed to remove some of the subjectivity surrounding segmentation scale selection. The combined approach for estimating scale delivered the best results when compared to the results obtained from individual metrics in both cases. Segmentation accuracy was significantly improved, although the spatial metric measuring undersegmentation did not perform as expected due to consistent selection of the smallest scale analysed in every trial. In both case studies, class delineation was significantly improved through use of the new segmentation metric. In addition, by-products from the results of the index calculation proved to contain valuable information on the occurrence of class gradients and areas of similarities that may be of great use in more natural environments such as lowland native grasslands.

4.1 Introduction

Object based image analysis (OBIA) approaches have become increasingly popular within the remote sensing community in the last two decades due to their strong performance and ability to provide high levels of classification accuracy (Blaschke *et al.*, 2014). Blaschke *et al.* (2014) even suggest that the development of these techniques constitute the formation of a new paradigm for image analysis. The shift from image pixels to image objects as the unit of analysis serves many advantages over traditional analysis approaches. Object-based approaches are able to reduce the potential rate of misclassification due to the generalising effects of segmentation (Costa *et al.*, 2014). The ultimate goal of image segmentation is to provide a sufficient level of generalisation so that image segments approach the optimal representation of their real-world counterparts, without overgeneralising or aggregating distinct features (Hay and Castilla, 2006; Blaschke, 2010). The variability of pixel values within a class is often high, especially in the case of high spatial resolution datasets. Within a per-pixel classification approach, the value to which individual pixels are compared during the classification process is typically a statistical variable derived from the total class data range. When classes encompass a wide range of data values, the variable chosen to represent a given class may not necessarily be a good representation of class properties (for example, a mean value for which there is a large standard deviation). In cases such as this, pixel values that fall outside the range of the class summary value are often misclassified, as their overall similarity to the class definition is lower than their similarity to an alternate class. Image segments, however, are typically also defined in terms of summary statistics. By using summary statistics to define the overall properties of a given segment, the total variability of pixel values within the segment is reduced, and the segment is more likely to approach the definition of the correct class.

In order to provide good classification results, all OBIA approaches require a good image segmentation (Hay and Castilla, 2006). During the segmentation stage, pixels are grouped together into homogenous regions, known as image objects, which are then used as the unit of analysis. The creation of these image objects serves many advantages, and can successfully mitigate issues associated with high intra-class variability (Blaschke, 2010). The size of image segments is primarily controlled by the segmentation scale factor, although some segmentation algorithms allow for the incorporation of additional factors determining segment attributes (Baatz and Schäpe, 2000). Good segmentations are characterised by segments exhibiting high degrees of internal homogeneity in conjunction with high dissimilarity between neighbouring objects (Blaschke *et al.*, 2014). The goal of segmentation is to produce an accurate representation of real-world objects, in order to facilitate accurate classification.

Identifying the scale factor at which image objects reach optimal levels of homogeneity, heterogeneity, and real-world correspondence has proved to be a difficult task. Formal assessment and determination of segmentation quality can serve to increase classification accuracy, however, the majority of currently proposed methods focus only on the assessment of spatial agreement between image segments and real-world objects (Costa *et al.*, 2014). In order to properly assess

the quality of an OBIA approach, we must determine if image objects created during the segmentation process are capable of representing real-world objects both in terms of class attributes and in terms of spatial and geometric properties (Zhan *et al.*, 2005). Many methods currently exist for the assessment of geometric accuracy within segmentations (Lucieer and Stein, 2002; Neubert *et al.*, 2006; Weidner, 2006; Clinton *et al.*, 2008), however, the majority of these approaches rely on empirical discrepancy methods designed to quantify the spatial and geometric disagreement between segments and reference objects (Neubert *et al.*, 2006). These approaches are limited, however, in that they only provide an assessment of geometric accuracy of the segmentation, but fail to account for variable user needs, and potential errors in the position and extent of reference objects (Costa *et al.*, 2014). In addition, methods relying solely on spatial agreement fail to provide information on the degree of similarity between reference objects and image segments in terms of class definitions (Whiteside *et al.*, 2011). The existence of spatial correspondence between reference and image segments does not guarantee good thematic correspondence. When image segments are unable to appropriately represent the thematic attributes of the classes to which they are to be assigned, then accurate classification is unlikely to occur. The determination of an optimal segmentation scale based on spatial agreement and geometric properties may not necessarily result in the identification of the optimal segmentation scale in a thematic context (Whiteside *et al*, 2011) and conversely, assessment from a thematic standpoint alone is unlikely to yield optimal spatial agreement. Therefore, the convergence of thematic agreement and spatial agreement must occur in order to accurately determine segmentation accuracy.

Figure 4.1 presents a series of examples at which thematic and geometric accuracies vary significantly with changing segmentation scale. If only the spatial agreement between the intersecting segment boundaries and the boundary of the reference object are considered, then the selected optimal scale may be incorrect in terms of accurate thematic representation of the targeted real-world object. Conversely, when only thematic accuracy is assessed, significant spatial errors may be incorporated into the final segmentation result, which again results in a sub-optimal representation of real-world objects in the final classification results. When both types of accuracy are considered, however, the segmentation scale that provides the best spatial representation of real-world objects, as well as the optimal thematic representation of classes, can be selected.

An additional issue with proposed segmentation assessment approaches is the failure to account for the occurrence of different spatial and thematic properties of classes. A single segmentation scale is often unable to accurately represent the full range of variability of different classes in terms of both spatial scale and thematic attributes, and therefore a multiscale approach is often needed. The wide variation in spatial and thematic properties between classes needs to be addressed in order to successfully predict optimal segmentation scale. Both types of agreement must be considered for each class independently within a segmentation in order for accurate classification and realistic representation of real-world objects to occur.

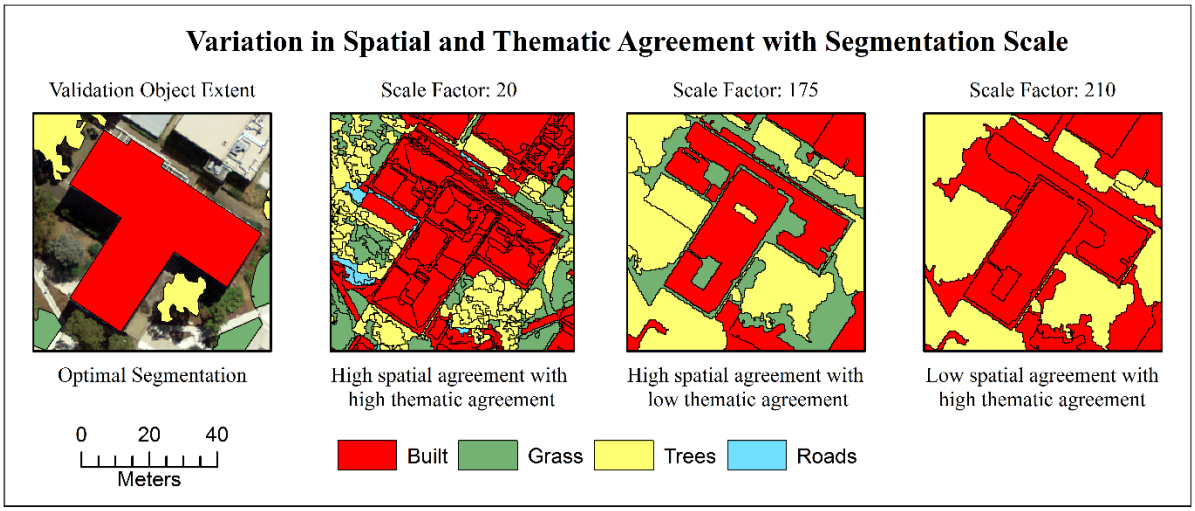

Figure 4.1: Variation in spatial and thematic agreement with spatial scale. The figure shows how selection of an optimal segmentation scale based solely on spatial or thematic agreement alone can result in poor representation of real-world objects, and that a unified approach is required for optimal results

4.1.1 Aims and Objectives

The purpose of this chapter was to provide a method for estimation of optimal segmentation scale for multiple thematic classes. As found in Chapter 2, the selection of a meaningful scale of analysis, and the subsequently selected image segmentation scale, is a vitally important component of any mapping approach. Selection of the scale at which communities achieve an appropriate level of generalisation is a key issue in the identification of ecologically meaningful remnant patches for many communities, not just lowland native grassland. Additionally, segmentation optimisation remains a very pertinent issue within the OBIA community at large, and development of a method capable of identifying such class-specific optimal scales represents a significant gap within the field.

The objectives of this chapter are therefore:

1. To develop and test a metric for assessment of segmentation quality that is capable of determining a class-specific optimal segmentation scale based on both thematic and spatial agreement
2. To employ the segmentation quality metric in different environments in order to determine the suitability of the approach for environmental applications

4.2 Methods

4.2.1 Case Study 1: Urban

A 15 cm spatial resolution aerial photograph was acquired over the University of Tasmania Sandy Bay campus in February 2011. The imagery consisted of four spectral bands, covering the blue, green, red and near infrared portions of the spectrum. The image was resampled to 30 cm spatial

resolution in order to improve computational efficiency. Figure 4.2 shows the dataset with associated reference objects. Two grey level co-occurrence matrix (GLCM) texture bands, homogeneity and second moment, were derived from the red spectral band using a 3x3 kernel as this band showed the greatest overall visual distinction between classes. Kernel size was minimised in order to avoid additional reduction in spatial resolution. Selection of the texture measures was determined via visual inspection for high levels of class differentiation. Texture was included in the analysis as it has been shown to be linked to many class characteristics and assist greatly in the differentiation of many land cover classes, particularly for vegetation (Wood *et al.*, 2012). Additionally, the inclusion of texture measures in image segmentation protocols has been shown to improve results (Laws, 1980). The resulting six band image was segmented in eCognition v 9 (Trimble Navigation, 2014) 55 times using scale factors ranging from 20 and 300, with an increase in the scale factor of five units for each segmentation. The shape and compactness criteria were set to 0.1 and 0.5 respectively, and all input variables given a weighting factor of one. A series of four thematic classes were selected from the image: buildings, grass, trees and roads. These classes were selected as they represent a wide variety of spatial scales, and cover the complete range of real-world objects present within the scene. A training dataset consisting of 73 polygons used to establish the definition of each class was digitised using ArcGIS 10.2 (Environmental Systems Research Institute, 2014). A validation dataset consisting of 100 reference polygons was also digitised in ArcGIS 10.2.

4.2.2 Case Study 2: Lowland Native Grasslands

As the selection of a meaningful scale of analysis was identified previously in Chapters 2 and 3 as one of the key issues in mapping lowland native grasslands, a case study was developed to test the proposed segmentation assessment method in such an environment. Tunbridge Township Lagoon is a small native reserve located near the town of Tunbridge in the Tasmanian Midlands. It was selected as the case study site for this chapter due to its small size, varied topography, and the presence of several types of key lowland native grassland communities. Additionally, the reserve is not formally grazed or fertilised as many similar vegetation patches in the region are, which allows for analysis of natural community characteristics.

In December of 2015, imagery was collected at Tunbridge Township Lagoon using a fixed wing unmanned aerial vehicle (UAV) equipped with a Sony α5100 (24 megapixels and 20 mm prime lens). Three red-green-blue (RGB) bands covering the visible portion of the spectrum were collected. The final spatial resolution of the output data set was 1.7 cm. Training objects were digitised for 3 thematic classes; *Acacia.dealbata, Lolium perenne* grassland, and *Themeda triandra* grassland. *Lolium perenne* is a commonly used pasture crop in the Tasmanian Midlands, and is typically found in disturbed areas of native grassland. It is of particular concern for lowland native grassland conservation given its vigorous seedling growth (Knox, Thompson and Campbell, 2006). Reference objects were digitised only for the *Acacia* and *Lolium* classes, as only these classes exhibited discrete community patches with clear object boundaries. The *Themeda* class

was excluded from the reference dataset, as it consists of a continuous connected area intergrading between the other classes. The *Themeda* community composes all portions of the scene not consisting of the remaining two classes, so the selection of optimal scale is less pertinent than for the more discrete classes. Image segmentation was performed 28 times, in increments of 10 for scale factor values between 20 and 300. Shape and compactness criteria were set to 0.1 and 0.5 respectively. It was determined that the red band of the image provided the best differentiation between classes it was given a weight of 2.0 in the segmentation, while the blue and green bands were given a weight of 1.0.

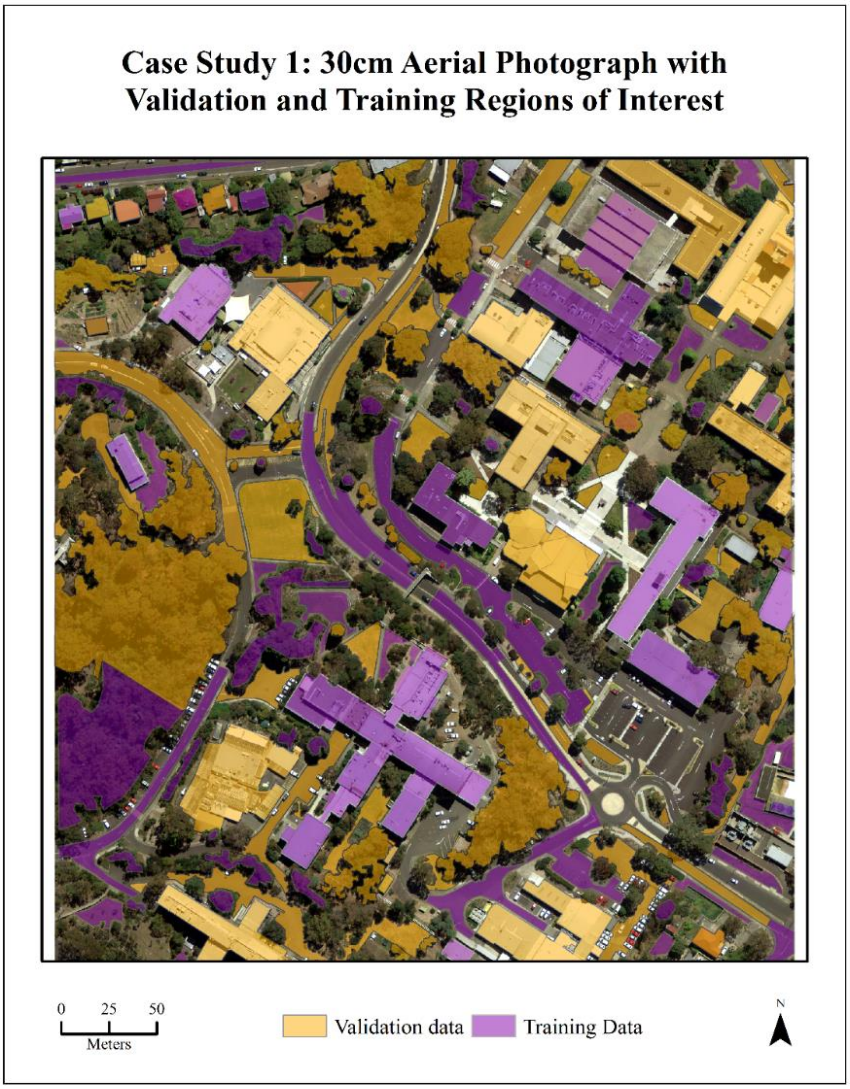

Figure 4.2: 30 cm aerial photograph of University of Tasmania campus, Sandy Bay, Tasmania,, with validation and training regions of interest. Digitised polygons show distribution of training and validation areas for the built, grass, tree and road classes.

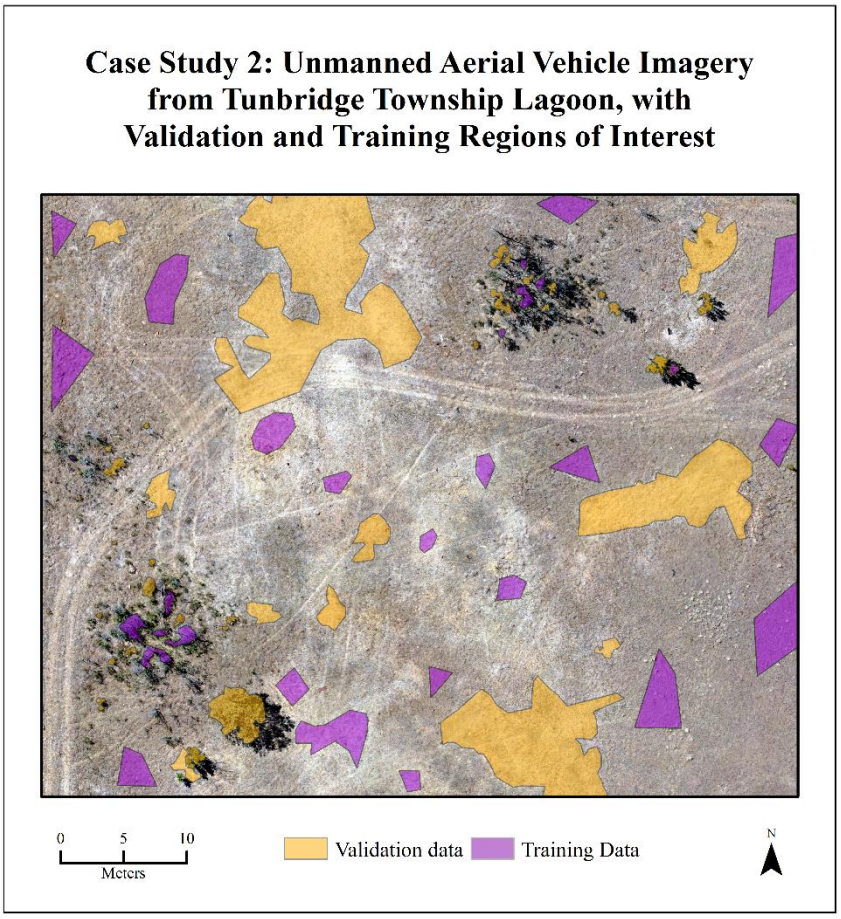

Figure 4.3 Unmanned aerial vehicle imagery from Tunbridge Township Lagoon, acquired at a spatial resolution of 1.7 cm. Digitised polygons show distribution of training and validation data for *Lolium perenne* and *Acacia* classes, in addition to training data created for thematic comparison with *Themeda triandra*.

4 .2.3 Calculation of Index values and Evaluation of Segmentation

Determination of Optimal Thematic Scale

Thematic accuracy within the context of GEOBIA can be defined as the degree to which image segments represent their real-world counterparts in terms of predefined class definitions. Class definitions are the range of image values, properties, or attributes said to be characteristic of a given thematic class. Such properties may include spectral values, texture measures, thematic attributes, or geometric attributes. As the segmentation scale changes, the number of pixels forming a given segment will also change. The addition and omission of pixels to image segments will alter the distribution of pixel values in terms of the input variables. These changes in the thematic and geometric properties of image segments result in varying degrees of segmentation optimisation. In order to determine the optimal segmentation scale, the scale factor at which image segments reach their most accurate configuration, in terms of thematic and geometric attributes, must be determined.

Class definitions are formulated based on reference objects or pixel values located throughout the scene. These objects are thematically homogeneous areas of pixels known to belong to a given class. The range of attribute values within the total set of reference objects for a class is assumed to be representative of the true thematic properties of the class, and used as the numerical class definition. Reference objects must also incorporate the geometric properties of the class, and represent a single image object that corresponds on a one-to-one basis with its real-world counterpart. The properties of each segment in a scene can be compared to the properties of each class, in order to determine the total degree of similarity. When the thematic properties of a segment match the definitions of a given class (i.e. the two distributions are sufficiently similar), the segment is said to be at its optimal thematic segmentation scale.

The optimal thematic segmentation scale for a given class will occur when two conditions are met:

1. The percentage of the total reference object area intersected by image segments with the greatest thematic similarity to an incorrect reference class approaches or reaches 0%
2. The distributions of attribute values within intersecting image and reference objects are at their highest possible similarity

The segmentation scale at which image segments intersecting the reference objects for a given class best approximate these conditions (i.e. achieve their highest values) can be defined as the optimal thematic segmentation scale. To determine at which segmentation scale these conditions are best met, an empirical evaluation of similarity between image segments and intersecting reference objects must be performed. Such an evaluation requires the use of a similarity metric. Similarity metrics are commonly used throughout the remote sensing literature as a means of calculating spectral separability of classes based on their unique characteristics (Van Aardt *et al.*, 2001; Tolpekin and Stein, 2009). As similarity metrics are able to determine the exact degree of agreement between the statistical distributions of two datasets, they provide a convenient and rigorous method of determining thematic agreement between classes and image segments. In order to accurately estimate the degree of thematic similarity between the desired image segments and their corresponding reference objects, the similarity metric selected in this experiment must exhibit the following characteristics:

1. The ability to incorporate multiple input variables

   It has been noted that many scale selection algorithms only assess image object attributes across a single input band (Yang *et al.*, 2014). As image classifications and segmentations are primarily performed on multi-band datasets, it is imperative that any evaluation of similarity between image objects and reference objects be able to account for variability across multiple input variables. Agreement across a larger number of input variables indicates a far stronger degree of thematic similarity, and a much higher probability of

correct classification, than agreement across a smaller number of variables. Similarity metrics must therefore be able to quantify the agreement between segments and reference objects in terms of empirical similarity within each input variable, in addition to providing an estimate of overall agreement across all input variables.

2. Facilitate the application of individual weighting factors to input variables

   Many commonly used segmentation algorithms allow for independent weighting of input variables (Baatz and Schäpe, 2000). It is important that the index used to evaluate thematic segmentation accuracy is capable of emulating such band weightings if they are applied. If different variable weightings are applied during the segmentation process, then the evaluation process must include these same weighting factors to variables in order to provide an accurate estimate of similarity.

3. Be able to evaluate non-normal distributions and be insensitive to statistical outliers

   The assumption of normality within data distributions is a common limitation of many statistical methods. Given the potentially large number of image objects and reference objects being compared, it is not feasible to test the distribution of every image object for normality. Because it often cannot be reasonably assumed that thematic classes are normally distributed across all variables, it is best to assume that the data does not conform to a normal distribution, and avoid metrics that have an inherent assumption of normality. Additionally, the small number of pixels forming segments at finer scales can create issues with many statistical models relying on normality, as there is often an assumption that the sample size meets minimum thresholds.

4. Provide rapid and computationally simple results

   Given the potentially large size of remotely sensed data sets (particularly high spatial resolution data, and hyperspectral imagery) the computation of exceedingly complicated metrics across multiple input variables for a large number of segmentations is simply unfeasible. A simple metric that is still able to provide accurate and reliable estimates of similarity is far more applicable in this instance, as it can provide results much more rapidly to researchers.

Based on these conditions, an index based on difference of means is inappropriate. Many metrics derived from the difference of means principle assume normality in the data, in addition to a minimum sample size. The tendency of the mean to be heavily influenced by outliers, particularly in cases with small sample sizes, is also of concern. In addition, many commonly used metrics are computationally complex, particularly over large datasets. Therefore, a simplified metric is proposed that utilises the variable median and quartiles to determine similarity. The median and

quartiles of a dataset provide superior representations of data distribution and characteristics under conditions of small sample size and non-normality, as both have been shown to be robust against outliers and do not inherently assume a symmetrical distribution of values around the median, as is often seen in mean-based metrics (Rousseeuw and Croux, 1993). This approach also provides comparison at three key points within the data distribution, thereby providing a more rigorous evaluation of similarity between datasets than a metric that uses a single evaluation criterion, such as difference of means.

The segment similarity metric S_{xvc} for segment *x*, class *c*, and variable *v* is defined as:

$$S_{xvc} = \sqrt{[(M_{xv} - M_{cv})^2 + (Q1_{xv} - Q1_{cv})^2 + (Q3_{xv} - Q3_{cv})^2]} \quad (1)$$

where M_{xv} is the median of segment *x* for variable *v*, M_{cv} is the median of reference class *c* for variable *v*, $Q1_{xv}$ is the first quartile of the segment, $Q1_{cv}$ is the first quartile of the reference class *c*, and *Q3* denotes the third quartile.

In situations where segmentation has been performed using multiple input variables to define object boundaries, where *V* is the total number of variables, the similarity between segment *x* and class *c* for the entire range of variables *v* in *V* is determined by the weighted mean of the similarity values. The similarity between *x* and *c* across the total set of input variables can therefore be defined as:

$$S_{xc} = \sum_{v}^{V} \frac{w_v S_{xvc}}{n} \quad (2)$$

where w_v is the weighting factor for variable *v* as defined by the segmentation parameters, and *n* is the number of segments intersecting reference objects belonging to class *c*.

In order to determine the segmentation scale at which the conditions for optimal thematic accuracy are best met, the similarity between each segment intersecting the reference dataset of a thematic class must be calculated for the entire range of intersecting segments (denoted as *X)* and the total number of classes, *C*. The class of maximum thematic similarity, $c_{max,}$ for segment *x* will be the class obtaining the lowest value of S_{xc} for all classes within *C*. As S_{xc} is calculated using the difference in values between the medians and quartiles of respective segments and reference objects, lower values indicate that there is less dissimilarity in the input distributions than for higher values. Condition 1 states that the scale of maximum thematic accuracy for a class will occur when the area of reference objects belonging to a given class composed of intersecting segments that have their $c_{max,}$ to an incorrect reference class approaches 0%. To evaluate the degree to which a segmentation scale meets this condition, the $c_{max,}$ for each segment *x* intersecting the set of reference objects belonging to *c*, must be compared. The degree to which Condition 1 is met for class *c* at a given segmentation scale can be represented by the percent of total reference object

area *A*, belonging to class *c*, that is intersected by segments with their c_{max} correctly identified as being *c*. This can be defined as:

$$Con1_c = 1 - \frac{\sum_r^R A(x_{corr} \cap x_{cr})}{\sum_r^R A(x_{cr})} \quad (3)$$

Where x_{cr} is a reference object with a class label of *c*, and x_{corr} is an image segment with its $c_{max,}$ to class *c* intersecting reference object x_{cr}

The degree to which a segmentation scale fulfils Condition 1 must be calculated for each thematic class composing *C*, for all segments in *X* intersecting reference objects in *R* in order to provide a value for $Con1_c$ for each thematic class at the given segmentation scale.

Condition 2 for optimal thematic scale states that the optimal thematic segmentation scale for a class occurs when the average value of S_{xc} for segments in *X* intersecting r_c where $c_{max,}$ of *x* is correctly identified as belonging to class *c* reaches its lowest value across the range of trialled segmentation scales (i.e values for S_{xc} approach zero). Therefore, for class *c* at a given segmentation scale:

$$Con2_c = \frac{\sum_r^R S_{xc}(x_{corr} \cap x_{cr})}{n} \quad (4)$$

Where *n* is the number of segments in $x_{corr} \cap x_{cr}$

As equation 2 requires the addition of agreement factors across multiple input variables, and the input variables used to determine the class definitions can possess different ranges, the data values across the scene needed to be scaled to the same range. Pixel values across both scenes were scaled between zero and one, using the entire data range of each variable with each respective scene. Equations 1 and 2 were implemented on both rescaled datasets independently in order to determine the degree of similarity between segments intersecting the various reference objects and each thematic class within each case study. The segmentation validation framework and equations described were implemented in Python 2.7. The agreement factor between each intersecting segment and each class was calculated iteratively using equations 1 and 2 in turn, so that each segment was assigned an agreement value for each class. $c_{max,}$ was determined for each intersecting segment within each segmentation scale by selecting the lowest value from the set of values determined by equation 2. $c_{max,}$ for each segment was then compared to the class of the reference object it intersected with at a given segmentation scale. For each reference object the area of intersection was calculated for each segment where $c_{max,}$ was correct (i.e. the same as the reference object that it intersects). Equation 3 was then implemented to determine the total percentage of reference object area composed of image segments with an incorrect $c_{max,}$. An overall percentage

was also calculated by determining the percentage area of intersecting segments with incorrect $c_{max,}$ for all reference objects within the scene. The degree of agreement was also calculated for each class within each segmentation scale by calculating the average index value for all segments with a correct class prediction (see equation 4).

Spatial Agreement

Many metrics have been developed to assess the geometric accuracy of an image segmentation, as summarised by (Whiteside *et al.*, 2011). It is generally agreed that undersegmentation has a much more detrimental effect on classification accuracy and subsequent representation of image objects than oversegmentation (Blaschke *et al.*, 2014). It is important that geometric accuracy is assessed independently of thematic accuracy in order to remove potential bias in comparison. To avoid additional complexity within the overall assessment, the degree of undersegmentation was calculated for each individual reference object (as originally formulated by (Weidner, 2008)) within the scene, and pooled across all reference objects within each class. Therefore, for class *c* defined by reference objects *R*, and being intersected by segments *X*, the degree of undersegmentation is said to be equal to:

$$U = \frac{X \cup R - X \cap R}{A(R)} \quad (5)$$

Where $X \cup R$ is the area of the union between *R* and *X*, and $X \cap R$ is the area of intersection between *R* and *X*, and A(R) is the total area of all objects within R

Combined Assessment

In order to fully assess a given segmentation scale, the final step is to combine the results of equations 3, 4, and 5 to create a metric that can evaluate conditions 1 and 2 for optimal thematic accuracy, as well as geometric accuracy, simultaneously. As the three metrics have slightly different ranges, in order to ensure equal contribution of each metric to the final index value, the produced per-class averages need to be re-scaled. This means that the segmentation scale achieving the highest value in a given metric will be assigned a new index value of 1, and the segmentation scale with the lowest value in the same metric will be assigned a new value of 0, with all other values being scaled proportionately in between. In some applications it may be desirable to weigh spatial agreement as of higher importance than thematic agreement or vice versa. For example, rule-based classifications in which classes are assigned primarily based on geometric object attributes such as shape, boundary relationships, or other spatial variables may necessitate the need for such weightings. In this case, each metric can be independently weighted to fit individual circumstance. Once scaling has been completed and weighting factors have been determined, the final index of segmentation quality S_q can be determined using the following equation:

$$S_q = \sqrt{[W_1 (Con1_c)^2 + W_2 (Con2_c)^2 + W_3 (U)^2]} \quad (6)$$

S_q was calculated for each of the thematic classes across all trialled segmentation scales in both case studies, as well as for the total set of reference objects in order to identify the best performing segmentation scale for the scene as a whole. The results for each class for the three individual metrics and the final combined metric were averaged to determine an alternate assessment of overall segmentation agreement.

4.3 Results

4.3.1 Case Study 1

Table 4.1 shows the optimal predicted segmentation scale for each class in the urban case study. $Con2_c$ typically predicts the optimal segmentation scale to occur at a much larger segmentation scale than the other two metrics, with the exception of the built class. The grass and road classes exhibit wide differences in their predicted optimal scales between their $Con1_c$ values and $Con2_c$ values. This discrepancy is much lower in the tree class, and for the total set of reference objects. When metric values are averaged across all classes, the optimal segmentation scale predicted by $Con2_c$ is significantly higher than the scales predicted by the other metrics, as well as those predicted for all of the thematic classes except the built class.

Figure 4.4 shows the change in metric values for the each class across the selected segmentation scales. In the built class, the sharp decrease in $Con1_c$ and Sq after scale factor 285 indicates the scale at which undersegmentation occurs, as larger segments are merged together. The change in $Con2_c$ from increasing to decreasing values around scale factor 90 shows the point at which the size of segments reaches the appropriate degree of generalization needed to approximate the class definition.

Unlike the built class, the behaviour of the road class in $Con2_c$ shows the highest degrees of similarity with mid-range segmentation scales. Smaller segments exhibiting the low spectral values typical of asphalt may bear increased similarity to areas of shaded vegetation found in between individual tree crowns and around the sides of buildings. This similarity may result in incorrect class predictions based on $Con2_c$. As the road surface generally exhibits a greater homogeneity of spectral values than these shaded regions, larger segments are easier to differentiate and assign correctly. The increase in $Con2_c$ is likely the result of undersegmentation Many of the larger roads in the scene contain dividers that are sufficiently spectrally different to reduce the degree of similarity between previously pure road segments and alternate classes.

The grass class exhibits the widest variation in $Con1_c$ of all the classes. The stability of $Con1_c$up to scale factor 110 is unusual, particularly given the rapid change in $Con2_c$ over the same range of scale factors. The rapid change in $Con2_c$ at lower scale factors indicates a large error of commission within this class, as a number of segments are being incorrectly predicted. The stability in $Con1_c$is also of interest, as it indicates that even though there is significant commission

within the class, the rate of omission is low. The consistently low values achieved for $Con1_c$between scale factors 20 and 110 indicate good delineation of small reference polygons within the segmentation. Aggregation of smaller grass patches results in the subsequent change in $Con1_c$ observed between scale factors 110 and 170. In terms of $Con2_c$ the treed class exhibits the poorest performance at finer segmentation scales. The most rapid decreases in $Con2_c$for the class, at scale factors 110 and 190, correspond to similar changes in $Con2_c$for the grass class. This indicates scales at which there is potential confusion between the two classes.

Table 4.1: Predicted optimal segmentation scale as determined individually by $Con1_c$, $Con2_c$, U, and S_q for each thematic class in case study one, in addition to scene-wide predicted optimal segmentation scales as determined by the total set of reference objects, and the mean agreement in each metric across all thematic classes

	$Con2_c$		$Con1_c$		U		S_q	
	Scale Factor	**Value**	**Scale Factor**	**Value**	**Scale Factor**	**Value**	**Scale Factor**	**Value**
Built	280	0.23	290	0.20	20	0.06	195	0.92
Grass	190	0.12	40	0.08	20	0.39	20	0.34
Trees	205	0.08	165	0.04	20	0.26	130	0.88
Road	190	0.06	35	0.17	20	0.09	55	0.35
Average	195	0.13	35	0.15	20	0.20	35	0.80
Overall	215	0.14	30	0.16	20	0.20	110	0.75

The final two panels in figure 4.4 show metric behaviour for the two methods of assessing segmentation across the entire scene. The approach derived from the total set of reference objects calculates values for $Con1_c$, $Con2_c$, and U by taking the mean calculated metric value of all segments with a correct class prediction across the entire scene. The average approach simply takes the mean value of the final class predictions within each metric. The approaches yield very similar results, however, there are some local variations, most notably between scale factors 60 and 70 where the total reference set approach predicts much more disagreement, which is reflected in $Con1_c$. The average based approach shows a much smoother change in metric values between segmentation scales than the alternate approach.

Figures 4.6, 4.7 and 4.8 show the predicted class extents for each of the four thematic classes and the overall scale predictions. As shown in figure 4.8 the predictions within the individual metrics for the two scene-wide predictions are similar. The total set of reference objects predicts a slightly larger scale factor for $Con2_c$, but the actual prediction of class extent is similar between the two scales. The overall scale prediction, however, as given by S_q, differs greatly between the two approaches. The averaged approach predicts a very fine segmentation scale as optimal, while the total set of reference objects predicts a much coarser segmentation. The actual class predictions

for image segments are again very similar, however, there is a significant reduction in oversegmentation with the prediction from the total set of reference objects.

Figure 4.4: Predicted optimal segmentation scale as determined individually by $Con1_c$, $Con2_c$, U, and Sq for each thematic class in case study one, in addition to scene-wide predicted optimal segmentation scales as determined by the total set of reference objects, and the mean agreement in each metric across all thematic classes

Figure 4.6 shows the metric results for the built and road classes. The predicted optimal scales for the built class in $Con1_c$ and $Con2_c$ are high compared to other classes. There is a notable degree of omission within the prediction from both metrics, however, there is very little observable commission. The prediction based on U presents an oversegmented scene. The areas of omission shown in the previous metric estimates are also present in the finer scale analysis here. The overall estimate provides similar predicted extent to the other results, but with slight improvements in the rate of commission. The scale prediction provided by $Con2_c$ for the road class shows significant

areas of omission, and a degree of undersegmentation along major arterial roadways. The estimates provided by $Con1_c$ and *U* provide much more complete predictions, however there is a tendency towards oversegmentation. The final scale prediction provides a good estimate, with very little omission, and good delineation of road boundaries. Dividers present in the main roads are excluded from the class prediction, and all streets are completely identified.

Figure 4.7 shows the predicted extents for the grass and tree classes. The scale factor of 190 predicted by $Con2_c$ for the grass class, shows a significant amount of commission, and many smaller patches are excluded from the predicted extent. The scale prediction provided by $Con1_c$ and *U* show much more detailed estimates, with fewer small patches being excluded. The final scale of 20 is identical to the prediction provided by *U*, and provides a good overall estimate. Small patches located between trees are clearly identified, and the degree of oversegmentation is relatively low. The tree class has good general performance at moderately high scale factors. As the class consists primarily of larger continuous stands of forest, this is not unexpected. The scale predictions for $Con1_c$ and $Con2_c$ however, have high degrees of omission, although commission is minor. The final scale prediction of 130 provides a reduction in the overall rate of omission, however there are still some areas that are omitted from the final predicted class extent.

4.3.2 Case Study 2

Figure 4.5 shows the changes in metric values for each class, the average of all classes, and the total set of reference objects. The most noticeable difference between the two classes is the smaller increase in *U* with scale for the *Lolium* class than for the *Acacia* class. As the *Acacia* class consists of a mixture of established trees and saplings in various stages of growth, this is to be expected. Both classes also exhibit a gradual decrease in $Con2_c$ with scale, although the *Acacia* class exhibits a sharp increase at the largest scales. Interestingly, there is a greater difference between the results obtained for the total set of reference objects and the average of both classes across all metrics than was observed in the previous case study. Here, the average approach and the approach based on the total set of reference object show different patterns of behaviour, specifically in $Con2_c$ at larger segmentation scales. Table 4.2 shows the achieved metric values for the *Acacia* and *Themeda* classes, and for the whole scene based on the reference objects of the previous two classes. The most noticeable difference between the performance of the *Acacia* and *Lolium* classes is the much higher values across all metrics for the *Acacia* class. The *Lolium* class performs well across all metrics, with consistently low values. The *Acacia* class however has very high values, particularly for $Con2_c$ and *U*. This indicates that there is significant disagreement between the segments predicted as belonging to the *Acacia* class and the training data, as well as a high rate of undersegmentation for the class. Figures 4.9 to 4.11 show the predicted class extents calculated for each class based on the individual assessment metrics and the combined metric for the second case study.

Figure 4.5: Calculated metric values for *Acacia dealbata* and *Lolium perenne* classes, as well as scene-wide predicted optimal segmentation scale as determined by the total set of reference objects, and mean value of the two input classes. Plot lines show changes in individual metric values for $Con1_{c,}$ $Con2_{c}$, UC_x, S_q and across the range of trialed segmentation scales

Table 4.2 : Predicted optimal segmentation scales for case study two in the three individual metrics and combined metric for *Acacia dealbata* and *Lolium perenne*

	$Con2_c$		$Con1_c$		U		S_q	
	Scale Factor	**Value**	**Scale Factor**	**Value**	**Scale Factor**	**Value**	**Scale Factor**	**Value**
Acacia	200	0.29	300	0.10	20	0.35	160	0.93
Lolium	230	0.05	110	0.07	20	0.01	120	0.64
Overall	300	0.07	210	0.08	20	0.14	120	0.74
Average	230	0.17	300	0.09	20	0.24	110	0.73

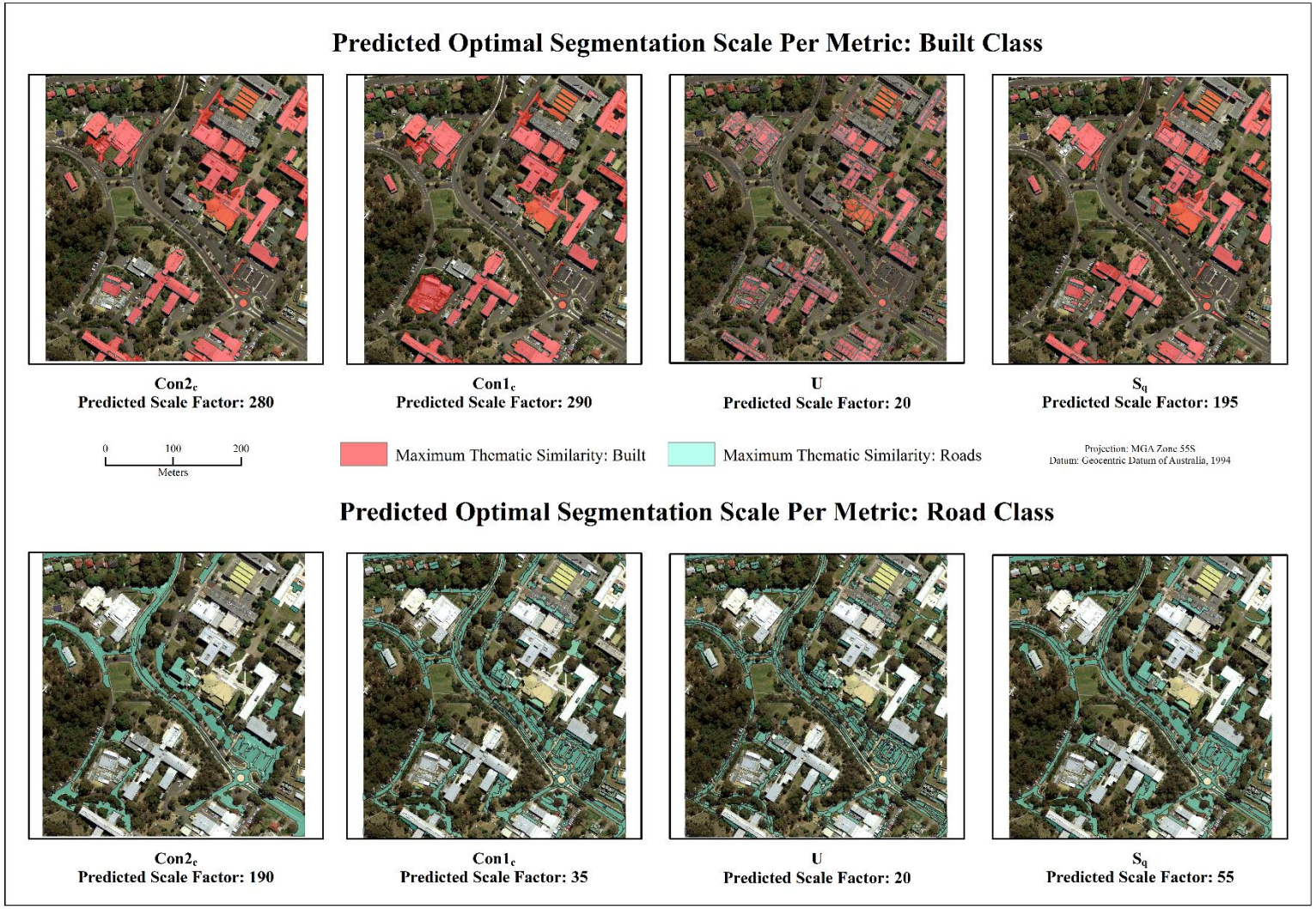

Figure 4.6 Predicted class extents for built and road classes based on the optimal segmentation scale parameter as predicted by $Con1_c$, $Con2_c$, *U*, and $\mathbf{S_q}$. Coloured areas show image objects determined to have their greatest thematic similarity to the given class at the predicted optimal segmentation scale as calculated by each metric.

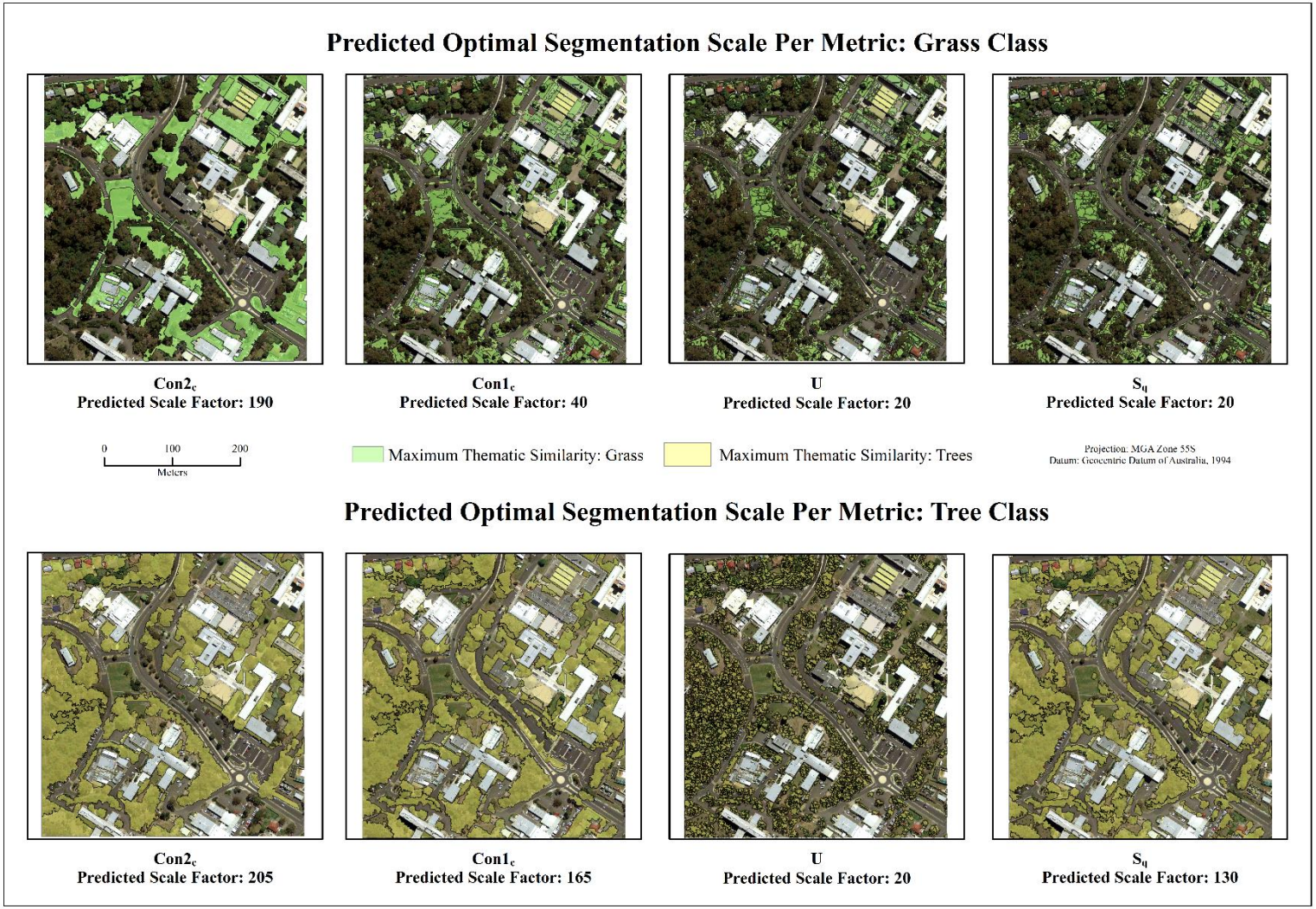

Figure 4.7 Predicted class extents for grass and tree classes based on the optimal segmentation scale parameter as predicted by $Con1_c$, $Con2_c$, U, and $\mathbf{S_q}$. Coloured areas show image objects determined to have their greatest thematic similarity to the given class at the predicted optimal segmentation scale as calculated by each metric.

Figure 4.8: Scene-wide prediction of class extent for all thematic classes based on the optimal segmentation scale parameter as predicted by$Con1_c$ $Con2_c$U, and $\mathbf{S_q}$. Coloured areas show image objects determined to have their greatest thematic similarity to each given class at the predicted optimal segmentation scale as calculated by each metric. The approach derived from the average metric values takes the average value of the results obtained from the four thematic classes to give an estimate of average class performance. The approach based on the total set of reference objects determines optimal segmentation scale by determining the metric value based on any image object with a correctly predicted thematic class of maximum agreement.

Figure 4.9 shows the predicted class extents for the *Acacia* and *Lolium* classes in $Con1_c$ $Con2_c$, *U* and S_q. For the *Acacia* class $Con1_c$, aprovides relatively complete prediction, although some smaller saplings are excluded due to the relatively coarse predicted scale factor of 200. There is some error of commission within the prediction, particularly for shaded regions around the base of larger established trees, however, there is no noticeable incorporation of grassland. The prediction from $Con2_c$ is similar to that of $Con1_c$, although there is an increase in the rate of omission, notably around the edge of the larger stands, particularly for dead or dry biomass. As seen in the previous case study, *U* predicts the optimal scale to be the smallest analysed, with a prediction of 20. This produces a largely overpredicted and oversegmented estimate of extent, with significant areas of grassland being incorporated. The combined prediction of scale factor 160 for the *Acacia* class provides a good, accurate estimate of class extent, with an improvement in both omission and commission errors from the estimates of $Con1_c$ and $Con2_c$. The prediction is complete and there is a minimal degree of oversegmentation.

The $Con2_c$ prediction for the *Lolium* class shows a series of large segments with a high degree of generalization. Areas of *Lolium* bordering with *Themeda* patches are frequently excluded, and there is also a noticeable amount of undersegmentation, particularly bordering the road. The prediction from $Con1_c$ of 110 provides much better delineation of patch boundaries. Confusion still exists between patches of *Lolium* and *Themeda* in transitional zones, however this is to be expected. As seen in the *Acacia* class, the prediction for U provides a poor estimate, with significant over-prediction of class extent an extreme level of oversegmentation.

Figure 4.10 shows the predicted optimal segmentation scale for the total set of reference objects and for the average of both classes. The average approach provides almost identical results to the predictions for the *Lolium* class alone, with the exception of $Con2_c$, which provides a much coarser predicted scale. This result is not unusual, however, as the *Acacia* class also has relatively good performance at the selected segmentation scales, as shown in figure 4.9. The predictions derived from the total set of reference objects are of interest, as they provide very different predictions than previously seen in the other classes. $Con1_c$ provides a much coarser predicted scale than seen previously, with a factor of 300. At this scale, both classes are significantly generalised, and there is a noticeable degree of undersegmentation. The prediction from $Con2_c$is presented as a midpoint between the coarse prediction from the *Acacia* class, and the comparatively low *Lolium* prediction. Performance by both classes is moderate, with a degree of omission being present in both. The *U* prediction is consistent again at 20. The final prediction of 120 is a good fit for both classes.

4.4 Discussion

4.4.1 Case Study 1

The performance of the built class based on $Con1_c$is very poor compared to the other classes. However, the values obtained for *U*, even at non-optimal segmentation scales, are the lowest of all

the classes. $Con1_c$also exhibits an unusually low increase in metric values with increasing scale parameter. The most likely explanation for the high values observed in $Con2_c$ within the built class is the amount of variability within the class. The variety of surfaces incorporated into the built class results in a wide range of spectral and textural properties, which results in a very broad class definition. The optimal scale prediction for the building class (shown in figure 4.6) identifies the majority of built objects within the scene, however, there are significant areas of omission. The major cause of omission within the segmentation is a high degree of similarity between the roofing material used in some of the larger buildings and the tarmac composing the majority of the roads within the scene. The rate of commission within the built class is very low, due to strong boundary delineation. The rate of oversegmentation is also very low, likely due to the homogeneity of roof surfaces, which greatly aids in the establishment of compact objects with well-defined boundaries The class extents predicted by each of the three incorporated metrics vary slightly from each other, however, the resulting prediction of the combined metric shows better delineation of built objects than the individual metrics. The predicted optimal segmentation scale does not introduce any additional errors of omission than from the individual metrics described previously. This indicates that the omission is likely a result of insufficient training data for certain types of buildings, or an inability to distinguish between the built and road classes within the input variables.

The remaining thematic classes have comparatively consistent values for $Con2_c$ across the range of tested segmentation scales. The values obtained for the tree class, and both the multi-class average and total set of reference objects all exhibit a gradual decrease in similarity with increasing scale. For the tree class, object boundaries are relatively well defined, although there is some degree of confusion with shaded areas. At smaller segmentation scales, delineation of individual tree crowns is excellent (as seen in Figure 4.7), and the amount of omission is low. However, the values obtained for $Con2_c$are at the upper end at these scales, as such small objects are unable to approximate the total class variance due to a high degree of oversegmentation. Although the class estimate based on $Con2_c$ does omit smaller trees from the predicted class extent, the rate of commission is low, and the overall extent of the class prediction remains relatively consistent with increasing scale. The predicted optimal scale factor of 130 presents a midpoint between the larger scale prediction based on thematic object similarity, and the smaller scale predicted by the spatial metric *U*. The final predicted scale factor provides an estimate at which omission rates and the degree of undersegmentation are balanced so as to provide the best relative prediction of the class. The scale prediction for the combined metric attempts to mitigate the different types of errors to similar levels so as to provide a balanced estimate of optimal segmentation scale.

The grass class has the smallest mean object size of the four thematic classes, with an overall optimal scale prediction of 20 (the smallest scale analysed). The degree of omission is similarly high at larger segmentation scales, particularly for smaller patches. The results obtained for $Con1_c$ and the combined metric are similar in their predicted extents, with very little observable omission or commission. The result obtained for $Con2_c$ however, shows a significant degree of

undersegmentation, with many grassed areas extending into the paved areas surrounding them. Given the variety in grass types and management practices, it can be expected that the grass class will have varying thematic properties across the scene, hence the high predicted scale factor for $Con2_c$ .Although the predicted optimal segmentation scale of 20 results in some oversegmentation, the predicted class extent is good, with smaller patches showing a strong geometric accuracy.

Figure 4.5 also displays the optimal segmentation scales for the road class. Of all the classes, the road class shows the most consitently low values for $Con2_c$ The predicted class extent determined by $Con2_c$ includes some buildings, and there is a significant amount of omission, but only a few small areas of commission appear in heavily shadowed areas. The results at scale factors 35 and 20 for $Con1_c$ and U are very similar; omission is low, and the areas of commission are confined to buildings with tarred roofs. The overall predicted class extent is good, with decent boundary delineation. There is a degree of oversegmentation within the class, at these particular scale factors, however, this is resolved in the final predicted scale factor of 55. The predicted extent is again very similar to the previously discussed scales, however, many oversegmented regions have been merged. Overall, the fit of image segments to the known class distribution is good. Object boundaries are well defined, and there is little to no extension of object boundaries into adjacent regions. The rail guard that separates the two lanes of the road traversing the centre of the scene has been excluded from the class predicted extent, indicating that the selected segmentation scale is significantly fine enough to exclude small and narrow features.

Figure 4.8 compares the predicted class extents for the optimal segmentation scale derived from the total set of reference objects, and from the average of the four thematic class. Overall class extent predictions are good, although there is significant generalisation in classes with smaller patch sizes such as the grass and road. The result obtained from the total set of reference objects contains less oversegmentation. Generalisation within the grass class is still extensive for the prediction based on $Con2_c$ however, this is in keeping with the overall trend towards higher degrees of thematic similarity at larger segmentation scales, even if spatial agreement is low. The two methods of overall assessment have returned almost identical results for U and $Con1_c$. The predicted scale factor is low for both, with the prediction U outside being the finest scale factor of 20.

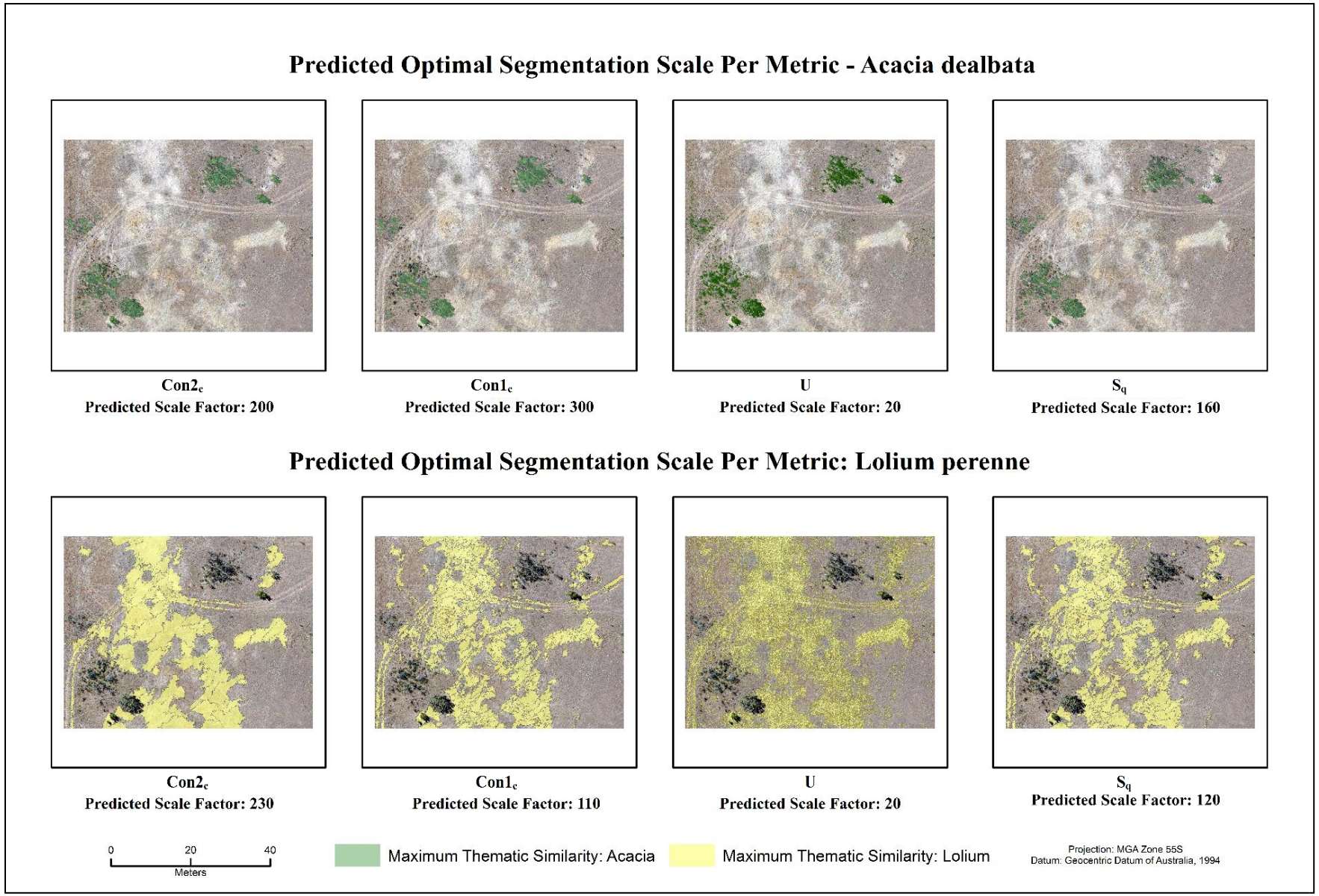

Figure 4.9: Predicted class extent for *Acacia dealbata* and *Lolium perenne* classes for $Con1_c$ $Con2_c$ UC_x, and $\mathbf{S_q}$ a. Shaded polygons indicate image objects that have their greatest thematic similarity at the given segmentation scale. Optimal segmentation scales are defined as the point at which each metric reaches its lowest value

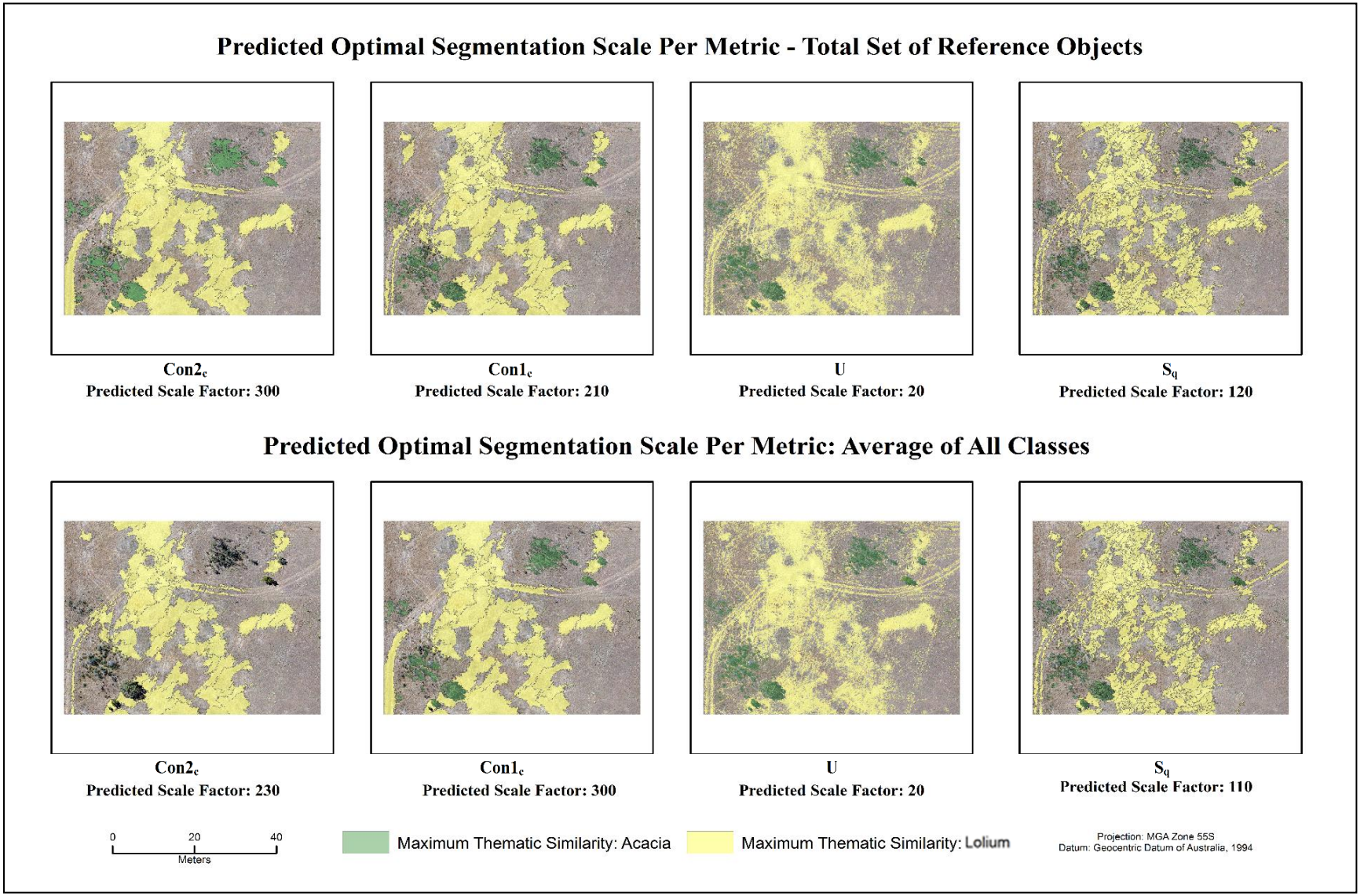

Figure 4.10: Predicted optimal segmentation scale factors for the two scene-wide assessment approaches, based on obtained metric values for $Con1_c$ $Con2_c$ UC_x, and S_q Shaded areas indicate polygons with their highest thematic similarity to *Lolium perenne* or *Acacia dealbata* .The scale predictions for the total set of reference objects are obtained based on the values calculated for image objects with their highest thematic agreement to the correct class, while the average approach takes the mean value of index results across both input classes.

Overall, there is a marked difference in the predicted optimal scale between the two approaches, with the averaged approach predicting a scale factor of 30, and the result from the total set of reference objects being 55. The prediction for the averaged approach is to be expected, as all four thematic classes have relatively good performance at smaller segmentation scales due to the high spatial resolution of the imagery. The high performance of individual classes at larger segmentation scales is mitigated by the poor performance of others, eventually leading to the prediction of a scale factor at which most classes perform moderately well. As can be seen in figure 4.8, the scale factor predicted by the total reference object approach produces a far better estimate of class extent. The predicted extent is good, with low degrees of omission and commission. Small patches are still identifiable, and larger patches are generalised appropriately. There is little to no confusion between classes that is not evident in the individual class predictions, and the fit of image segments to reference objects is good both in terms of boundary delineation, and the number of segments composing each object. The prediction derived from the averaged approach produces a highly oversegmented scene, with no improvement in thematic accuracy. The areas of omissions and commission for each class are essentially identical to those in the prediction from the alternate approach.

4.4.2 Case Study 2

Case study 2 shows the performance of the assessment approach in a less structured environment. The communities analysed in this example are known to exist along continuums, and are much more difficult to discretely compartmentalise into homogeneous image segments compared to the classes in Case Study 1. The segmentation validation approach proposed here, however, can still provide valuable insight into class properties and scales of occurrence.

As Figures 4.9 and 4.10 show, the individual metrics alone struggle consistently to provide good estimates of optimal scale. $Con2_c$ has predicted scale factors that are far too coarse for the selected classes, resulting in a loss of class boundaries and significant undersegmentation. The *Acacia* class has similar issues at the scale factor predicted by $Con1_c$, where there is poor prediction for smaller saplings, and omission of dead biomass. There is also significant confusion with shadows, resulting in object boundaries being misplaced. The predicted extent for the *Acacia* class also shows many gaps where areas of tree crown are omitted from the predicted extent. There is also a noticeable trend in omission of trees growing on the north-western edge of the larger stands, with object boundaries being placed further to the south-east in order to capture shadows. Additionally, the values for $Con2_c$ are significantly larger than the values obtained for the *Lolium* class across all segmentation scales. This is likely due to the consistent inclusion of shaded areas in the predicted class extents, but not in the training data. The prediction based on *U* is very poor, and shows extensive over prediction of class extent. Objects scattered throughout the study site, such as car parts and rusted iron, are also identified as *Acacia* in this result, which only increases the rate of over prediction.

The *Lolium* class has overall better performance than the *Acacia* class, although it still has a very coarse scale prediction based on $Con2_c$ Similar to the *Acacia* class, the results for U are poor, with significant over prediction of extent, and both significant omission and commission. The results obtained for $Con1_c$ are very close to the predicted optimal scale factor, and present

a relatively good overall prediction. In the final results, there is some commission and omission around object boundaries, due to the natural gradients existing between communities. The identified patches of *Lolium* do however form relatively homogeneous patches, and overall patch boundary delineation is good at the overall predicted optimal segmentation scale.

For this case study, the overall predicted optimal segmentation scales are very similar between the two scene-wide approaches, despite the difference in individual metric prediction. Interestingly, the predicted scale factors based on $Con1_c$and $Con2_c$ are almost reversed between the two approaches. Both results produce a highly undersegmented result, with significant omission and commission in both classes. The prediction of a high scale factor for $Con1_c$is primarily driven by the good performance of the *Lolium* class at larger scale factors in this metric. The *Lolium* class is far more expansive than the *Acacia* class, which is reflected in the reference dataset, in which the reference objects are also much larger. Additionally, the metric values obtained for the *Lolium* class are generally lower than those attained for the *Acacia* class, as discussed previously. As this approach to overall scale prediction works by identifying the scale factor at which the lowest similarity metric value is achieved across all image segments with a correct class assignment, the difference in sample size appears to have heavily biased the optimal scale prediction towards the properties of the *Lolium* class. The prediction for $Con1_c$ is affected in much the same manner as the prediction for $Con2_c$. Performance in the *Acacia* class is comparatively poor at a scale factor of 230, however, the good performance of the *Lolium* class at this scale factor essentially overrides such performance issues. The overall scale prediction of 110 is a good fit for both classes, with both exhibiting favourable performance in the combined metric.

The approach based on the average performance of the thematic classes provides an interesting counterpoint to the results from the total set of reference objects. In this case, performance of both classes is weighted identically, even though the extents of the classes vary significantly. The final scale prediction is very similar to the result from the previous approach, and provides a good overall fit. The predicted scale factors for $Con1_c$ and $Con2_c$ much better optimize the performance of both classes, as it can be clearly seen in figure 4.10 that the comparative performance of both classes is much closer than in the scale predictions for the total set of reference objects.

As values for $Con2_c$are calculated for all image segments within the scene, and for all thematic classes, the values can be mapped in order to show potential areas of confusion and to aid in the identification and quantification of community gradients. As $Con2_c$ values are calculated by comparing the distribution of values within a training data set and individual segments, values can be calculated for the *Themeda triandra* class in addition to the *Lolium* and *Acacia* classes. Training datasets for the purpose of calculating $Con2_c$ and subsequent training of a classifier do not need to have the discrete object properties required of the reference dataset which is used to calculate $Con1_c$ and *U*. These metrics essentially validate the similarity prediction derived from the training data. Figure 4.11 shows $Con2_c$ values obtained for each image segment for both the *Lolium* class and for *Themeda triandra.* Values for

$Con2_c$ approaching 0 indicate a good match between the image segment distribution and the distribution of the target class, while values approaching 1 (one) indicate the maximum potential disagreement between the two. From the figure, it is evident that similarity values between the two are complementary, in that where there is high similarity to one class, similarity to the other is low. For both the *Lolium* and *Themeda* classes, it is evident that as distance from the centre of the patch increases, so too does disagreement with the class definition. There is very little indication of significant confusion between the two class types, with no segments being identified as having significant similarity to both thematic classes. The segments which belong to the *Acacia* class show a significantly higher degree of similarity to the native *Themeda* class than to the introduced *Lolium* class.

4.4.3 Metric Performance

Overall, the combined metric provides good predictions of optimal segmentation scale, and for the scene as a whole. Predicted scales and extents for classes in both case studies are good, with the combined metric providing a genenerally better prediction in terms of the two types of agreement assessed than the individual metrics alone. Although the final predictions contain some discrepencies in terms of both thematic and geometric fidelity between image segments and reference objects, the predictions still provide an empirically justifiable prediction of optimal segmentation scale for each class. It is important to remember that optimal segmentation scales in terms of thematic attributes and spatial attributes are unlikely to occur at the same spatial scale, as segmentation scales with high thematic agreement tend to be larger than those with high spatial agreement. This is due to the fact that an increase in the total number of pixels within a segment usually leads to better representativeness of class characteristics, and a reduced effect of outlier values. Converseley, high spatial agreement tends to favour segmentation scales with smaller average object sizes, as even if intersecting segments are located predominantly outside a given reference area, their small size results in a much lower degree of undersegmentation. The aim of the combined metric is to determine the point at which high spatial and thematic accuracy occur together, while the individual metrics provide estimates of thematic and spatial agreement in isolation. The predicted overall optimal segmentation scale for a class is essentially the point at which the two types of agreement are optimised relative to one another, though not necessarily optimised individually.

The similarity metric $Con2_c$used here to quantify thematic distances has several advantages, such as the fact that input variables can be individually weighted to mirror the weighting of variables within the segmentation or classification portions of the final analysis. This metric does not, however, evaluate the correctness of class assignments within any form of spatial context. The similarity metric only provides an estimate of similarity between segments and classes, and provides no consideration of the total percent of reference area correctly identified. Used alone, this metric can only approximate thematic agreement outside of the spatial realm. The overall framework of the methodology, however, is able to accommodate changes in the specific algorithm used to evaluate similarity. It is entirely possible for a user to use an alternate similarity measure as long as a value of agreement can be determined for each input class, and scaled appropriately. This gives the approach an additional degree of flexibility in cases where the similarity metric proposed here may not be appropriate for the selected data or application.

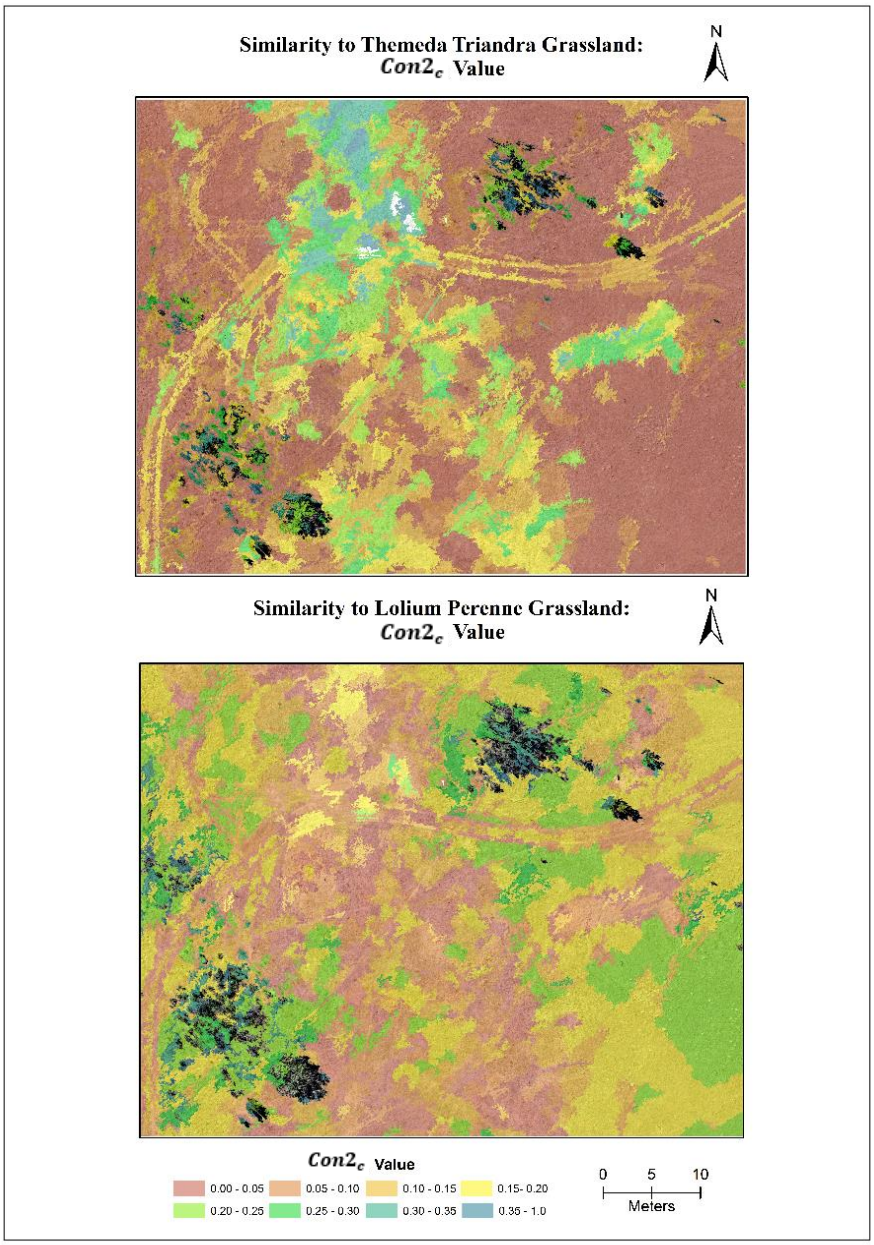

Figure 4.11: Per-object $Con2_c$ values to both *Themeda triandra* and *Lolium perenne* at the optimal predicted scale factor for the *Lolium* class of 120. Red objects indicate a high degree of thematic similarity, while green and blue values indicate disagreement. As can be seen below, the overall distribution of similarity values is complementary between the two classes

$Con1_c$ is designed to determine the segmentation scale at which the lowest percentage of total reference object area is intersected by segments with their highest thematic agreement to an incorrect thematic class. This metric is useful as it provides an idea of how well reference objects and their associated class are represented spatially by the image segmentation. This metric does not provide a description of how good the match between objects and the relevent class definition is; it provides a mid-point between estimates of spatial and thematic accuracy by predicting the correct area of reference objects for a given class and attempting to optimise

it. Essentially, this metric takes the results of $Con2_c$ and applies them within a spatial context. This metric does not, however, account for the level of under or over segmentation within a reference object. It is entirely possible that although the area of intersection is correctly identified in its entirety that the selected segmentation scale may present significant spatial error in terms of misplacement in segment boundaries. The combined performance of $Con1_c$ and $Con2_c$ provides a strong indication of optimal thematic segmentation scale, although there is a slight tendency towards undersegmentation in the results

In terms of individual metric performance, the spatial assessment metric,*U* did not contribute to the final results as signficantly as hoped. Ideally, the metric would predict the scale factor at which there is a one-to-one correspondence between image segments and reference objects in terms of boundary placement and spatial properties, as seen in the first panel of Figure 4.1. However, the metric instead considered the smallest analysed segmentation scale as optimal. The likely reason for a failure of the metric to produce the desired results is variability in the size of reference segments. As many of the classes examined in both case studies contain objects of varying size, it can be difficult to identify a scale factor at which every single reference object reaches one-to-one correspondence, as the object will not reach such a state at a single scale. If the analysed classes contained uniformly sized discrete objects, then perhaps this approach would have predicted alternate optimal scales, however, this condition simply cannot be met in the datasets used here. The spatial index did, however, provide a useful balance in the final metric against the coarser predictions of $Con2_c$ which tended towards undersegmentation in its predicted optimal scale factors. The consideration of spatial agreement is still an important factor in the determination of overall optimal segmentation scale, and should be included in future improved assessment approaches. The inclusion of an alternate assessment metric will likely provide results better aimed at guiding optimal scale selection from a geometric and thematic standpoint.

The combined approach of using three individual metrics to determine overall class-specific optimal segmentation scale has several advantages, both within the indivual components and within the method as a whole. The similarity metric, designed to determine the level of disagreement between image segments and each thematic class, allows for empirical evaluation of overall thematic similarity. It can also provide a means for determining the suitability of both the training and validation data to be used in the final classification process. If the results of the similarity analysis are consistently low within a given class, this is indicative of a potential mismatch or misrepresentation of class properties between the associated training and validation datasets. Conversely, if the returned values for objects with predicted highest thematic similarity values to a given class are consistently low, a high degree of similarity between the training data and validation data can be assured. Additionally, values are calculated for each segment and each thematic class independently, which can allow for comparision of similarity values between classes, which can be a powerful tool for later image analysis, as seen in Figure 4.11. This may be of use in cases where there are errors in the predicted class extent, or when one or more classes have overlapping or non-distinct class definitions. The potential for individual class similarity values to be analysed in further steps provides a means of numerically quantifying the potential for later classification errors.

The results of the case studies presented here indicate that per-class prediction of optimal segmentation scale is possible. By assessing individual class properties, the approach has produced strong, indepent prediction of class-specific optimal segmentation scale. Both case studies highlight the importance of class-specific scale analysis, and how generalised scene-wide approaches can result in poor class representation in final segmentations. The approach illustrated here can predict overall optimal segmentation scale with some success, however, the individual class-specific results provide a much better overall representation of scene features, and should strongly be considered over the scene-wide result. The results of the two scene-wide approaches performed very differently in the two case studies. For the first case study, the results for the total set of reference objects approach provided signficantly better results than the average approach. As class extent and scale of occurrence is relatively balanced between the four classes analysed in this case study, the approach using the total set of reference objects is much better able to determine the optimal segmentation scale, as it considers the best fit for the reference objects as a unit. However in the second case study, the averaged approach gave a better approximation, due to the discrepancy in the size of the reference datasets between the two classes. The choice of which approach to use to determine overall optimal segmentation scale therefore will depend on the properties of the analysed classes. If there is a significant imbalance in the extent of thematic classes which is reflected in the reference data, then the average approach will yield better results for all classes, particularly if there is a big difference in the mean size of objects per class, as seen in with the individual *Acacia* crowns and the larger patches of *Lolium* in the second case study. As the class results are averaged for all of the input classes, the overall weighting of individual class performance will be equal, and the dominance of one class will be mitigated. In examples such as the first case study where there is a balanced representation of classes across the scene, then the total set of reference objects approach will provide better overall scale prediction, as one class will not dominate the results.

The methods employed in this chapter focus on the determination of single optimal segmentation scales for individual classes, and apply the indices within a single hierachical level. One of the foundational strengths of OBIA approaches is the ability to create vertical hierachical relationships between image objects within sub and super levels. OBIA is able to exploit these relationships at varying levels within this hierachy as a means of incorporating user knowledge. Varying super and sub-objects within a class hierachy may occur at different spatial scales throughout the classification hierachy. Iteratively segmenting and merging objects to create new levels above and within the hierachy can be a powerful tool in the refinement of classifications within this framework. The methods proposed in this chaper are not designed to be a definitive method for the determination of optimal segmentation scales, and are not designed to be used at the exclusion of user knowledge and contextual information. Rather, the methods described here can be used in conjunction with such approaches, and provide confirmation of choices made by users during the segmentation process. This method can also be applied across multiple levels within the hierachy to predict the varying scales at which class performance may be optimised.The prediction methods described here can also be used as a starting point within the analysis process, and provide valuable information about the information contained within the reference datasets and the image itself. As these methods rely

only on numerical comparison of segments to reference classes, there will always be a degree of error and uncertainty within the predicted outputs. User knowledge and contextual variables can act as a vital refinement process to correct for any errors within the predicted scale factors, and can be used to further improve upon the segmentation results.

An important final caveat of this methodology is that this method does not fulfill the same function as a formal image classification. The goal of the metric is to serve as a guide towards selecting segmentation parameters at which both thematic and spatial accuracy are high, and comparatively optimised. The results of the various metrics employed here cannot inherently assure good classification results, as final classification results depend heavily on the type of classifier used and the quality of the input data. However, the approach can ensure that each selected segmentation scale matches the thematic and spatial attributes of the class it is to represent as best as possible given the input data. This in itself may serve to increase the likelihood of a good classification, as classification accuracy has been shown to be heavily determined by segmentation quality (Hay and Castilla, 2006; Blaschke, 2010). This approach is also somewhat limited in that there is a requirement for validation reference objects to have clearly identifiable boundaries in order to assess geometric performance. In cases where not all classes exhibit such characteristics, then concessions can be made, as seen in case study 2 in which the *Themeda* class was not formally assessed. In highly heterogenous natural environments in which all target classes exist along gradients, this approach may not be appropriate. However, the use of $Con2_c$ alone, as seen in Figure 4.11, may provide key insights into community gradients, and aid in the identification of potential discrete boundaries that may not be immediately visible in the raw input data.

4.5 Conclusions

This chapter outlines a novel approach for the prediction of optimal segmentation scale for multiple thematic classes. The approach uses a combination of three metrics to assess both thematic and geometric accuracy within image segmentations, and provides the first approach for segmenation assessment capable of predicting optimal segmentation scale across multiple input variables on a per-class basis. Generally, the metrics designed to assess thematic agreement, $Con1_c$and $Con2_c$ performed much better than the metric assessing spatial agreement, *U*. Individual metrics failed to provide accurate predictions of optimal segmentation scale capable of meeting the required conditions for accurate segmentation. The combined metric, however, provided good prediction of optimal segmentation scale across all thematic classes employed in both case studies. Predicted segmentation scales were able to reduce errors of omission and comission, as well as reduce undersegmentation within the predicted class extents. Although there are errors within the predicted class extents for the overall optimal segmentation scales, this likely is the result of insufficient training data, rather than an issue within the approach itself. Scene-wide optimal segmentation scale was also predicted using two differing approaches. It was found that scene-wide scale prediction did not produce as strong results as the individual class-specific results, with many classes within the predicted extents performing poorly. This finding serves to highlight the need for class-specific scale assessment in order to accuractely portray class characteristics within a segmentation.

The case studies presented in this chapter show two relative extremes in terms of object properties. The first case study, consisting of an urban environment, contained thematic classes with discrete, well defined image objects, while the second case study showed a naturalised environment with less clearly defined object boundaries. The results of the two case studies indicate that the assessment approach is capable of providing estimates of optimal segmentation scale in both types of environments. In order to calculate $Con1_c$ and U however, reference objects must be discrete and have clearly identifiable boundaries within the imagery. This means that in cases where classes exist along continuums, such as the *Themeda triandra* class shown in case study 2, these metrics cannot be calculated properly. Although not ideal, this approach means that certain classes can still be optimised within the segmentation, and results may still well be improved with the final classification results. Additional outputs of the segmentation assessment process, such as maps of varying $Con2_c$ values across the scene, as seen in Figure 4.11 are invaluable in cases such as these, as they can be used to identify and quantify community gradients, and potentially aid in the delineation of reference segment boundaries that may not have been obvious in the original imagery.

4.6 Thesis Context

This chapter provides a solution to the issue of determining the optimal scale of analysis within OBIA approaches, as outlined in objective 2. The developed method of assessment is presented in the form of two case studies in order to determine the utility of the approach in different types of environment. The development of this segmentation validation framework is an important step in mitigating issues identified in chapter one, particularly surrounding the small patch size often associated with lowland native grassland communities and the spatial co-occurrence of communities. Case study two indicates that there is strong potential and utility for the approach, and that accuracy can be significantly improved for classes when this approach is employed. The results of Figure 4.11 also indicate that the approach can be used to identify community gradients, and potentially aid in the delineation of reference object boundaries in cases where boundary identification may be difficult. As native vegetation communities within the Midlands occur at widely differing spatial scales, there is often variation of patch size within single communities, and this approach provides a much needed degree of empirical validation to the scale selection process for each class.

Chapter 5

High spectral and spatial resolution imagery collected with an unmanned aircraft system (UAS) for analysis of lowland native grassland communities in the Tasmanian Midlands

Abstract

This chapter presents the results of a study undertaken to classify lowland native grassland communities in the Tasmanian Midlands region using high spatial and spectral resolution data captured using an Unmanned Aircraft System (UAS). Data was collected using a 20 band hyperspectral snapshot sensor with a spectral range of 600-875 nm. Four vegetation communities were identified for analysis in keeping with the four-class configuration used in Chapter 3. In addition to the hyperspectral UAS dataset, a Digital Surface Model (DSM) was derived using structure-from-motion (SfM). Training data was created using a series of points derived from fieldwork undertaken in November and December of 2015. Reference objects were digitised for each of the four classes, and two 100 m transects were used for secondary validation. Classification was undertaken using an object-based approach, and segmentation quality was assessed using the methods developed in Chapter 4. Optimal segmentation scale factors were determined for each of the four vegetation classes, in addition to the scene as a whole. Class-specific optimal segmentation scales ranged from 250-500, while the scene-wide predicted optimal scale factors were equal to 1,700 and 400 for the two segmentation accuracy assessments, as described in Chapter 4. Image segmentation was performed using the class-specific scale predictions to create a single image segmentation that was used as the basis for image classification.

Classification was undertaken using a random forest model (RF). Internal cross-validation accuracy from the training model was 97.4%. Variable importance measures from the training model indicated that the DSM was the most significant variable. Key spectral variables included bands two (620.9 nm), four (651.1 nm) and 11 (763.2 nm) from the hyperspectral UAS imagery. Classification validation was performed using both the reference segments and the two transects. For the reference object validation, mean accuracies were between 70 and 72%. Significant confusion between the *Danthonia* and *Wilsonia* classes resulted in poor accuracies for the *Danthonia* class, ranging between 37% and 46%. Classification accuracies based on the validation transects achieved a maximum overall classification accuracy of 93%. Class-specific accuracies based on the validation transects were higher in all three classification results.

The segmentation assessment method performed well, with predicted class extents conforming to known community distributions. The findings re-iterate the need for high spatial and spectral resolution datasets in order to accurately identify and assess lowland native grassland communities.

1 Introduction

The application of Unmanned Aircraft Systems (UAS) for environmental remote sensing applications has become increasingly prevalent in recent years. The ability of UAS to provide ultra-high spatial resolution datasets at a relatively low cost makes them an attractive option for many researchers (Anderson *et al.*, 2013). The development of commercially accessible platforms has led to a rapid increase in the applications for which UAS have been used. The applicability of UAS to grassland monitoring and mapping is particularly attractive due to the ability of such systems to collect spatially detailed datasets on demand. This ability is integral to grassland remote sensing due to the high seasonal variability observed in communities (Goetz, 1997; Wen *et al.*, 2010; Xu *et al.*, 2013).

Several studies have employed UAS as the principal platform in grassland research (Laliberte *et al.*, 2007; Rango *et al.*, 2009; Lu *et al.*, 2016). Although applications are primarily focussed on small-scale studies of agricultural productivity, such as estimating biomass (Kawamura *et al.*, 2011), several studies have focussed on broader-scale ecological applications of UAS for various applications within grassland environments such as monitoring degradation and change (Svoray *et al.*, 2011), mapping species regeneration post-fire (Silva *et al.*, 2014), estimating ground cover in rangelands (Breckenridge *et al.*, 2016), identification of grassland vegetation (Burai *et al.*, 2015), and assessment of species composition (Lu *et al.*, 2016). The most prevalent area of grassland research using UAS, however, is for rangeland monitoring and mapping. Extensive work has been undertaken, particularly in the south-western United States, to determine the feasibility of UAS for broad-scale, high spatial resolution analysis of semi-arid grassland and shrub communities (Laliberte *et al.*, 2007; Rango *et al.*, 2009; Laliberte *et al.*, 2010, 2011).

The majority of remote sensing studies using UAS within the realm of ecological research have focussed on the use of ultra-high spatial resolution datasets collected using broadband multispectral sensors or RGB cameras due to their low cost (Anderson *et al.*, 2013). The use of broad band multispectral sensors is not always capable of providing sufficient spectral detail for accurate analysis of vegetation types and attributes, even when data is acquired at high spatial resolutions. Applications of hyperspectral sensors using UAS platforms are still limited, and few studies have investigated their application for precision agriculture (Zarco-Tejada *et al.*, 2005; Berni *et al.*, 2009; Zarco-Tejada *et al.*, 2012; Zarco-tejada *et al.*, 2013; Aasen *et al.*, 2014; Gevaert *et al.*, 2015). Due to the fact that the majority of previously available high spectral resolution sensors are based on pushbroom designs, high fidelity Global Navigation Satellite System (GNSS) and Inertial Measurement Unit data were required for the creation of useable outputs (Aasen *et al.*, 2014; Aasen *et al.*, 2015). This issue has led to limited use and application of UAS mounted hyperspectral sensors within the ecological remote sensing community. The development of frame-based and snapshot hyperspectral cameras, however, eliminates the need for complicated geometric processing, and makes the collection of hyperspectral datasets from UAS much more feasible. The use of such sensors has enormous potential for ecological vegetation mapping and monitoring due to the high degree of spectral information captured.

1.1 Aims and Objectives

This chapter aims to assess the applicability of high spectral and spatial resolution UAS imagery to lowland native grassland mapping. By employing a high spectral and spatial resolution approach to community classification, the issues identified in previous chapters can be addressed and overcome. Additionally, this chapter serves as a case study showing the applicability of frame-based hyperspectral sensors for ecological mapping and monitoring.

2 Methods

2.1 Study site and vegetation communities

In November of 2015, imagery was collected at Tunbridge Township Lagoon, in the Tasmanian Midlands. The town of Tunbridge is located between the two major settlements of Hobart and Launceston, and marks the divide between the Northern and Southern Midlands regions. The lagoon serves as the only formally protected lowland native grassland habitat in Tasmania, and contains important remnant patches and many endangered species. The reserve covers an area of approximately 16 hectares, and has a wide floristic diversity. The western third of the site is populated by remnant *Themeda triandra* grassland, interspersed with *Acacia dealbata* and *Bursaria spinosa*. This portion of the site is steeply sloped in an easterly aspect. The remaining two thirds of the site are predominantly flat, and covered with a saltwater lagoon. The saltpan surrounding the lagoon is populated by many saline tolerant ground cover species, such as *Wilsonia rotundifolia, Sellieria radicans,* and in places the Australian Saltmarsh grass *Puccinellia stricta.* The areas between the saltpan and the bounding western and southern fences are populated by remnant *Danthonia trenuior* and *Poa labillardierie* grasslands. Vegetation communities are generally in good condition, although the southern side of the lagoon and a small area at the foot of the hill immediately adjacent to the lagoon is still recovering from unplanned burning in the summer of 2014.

For the purpose of this study, a subset of the total reserve area was targeted. This area is found on the south-western corner of the lagoon, and covers a transitional area between saltmarsh vegetation, native grassland communities dominated by *Danthonia trenuior* or *Poa labillardierei,* and the foot of the hill which is dominated by the *Themeda triandra* community. A total of four vegetation classes were identified for analysis, as well as a soil class. The first class consists of the saline vegetation communities found surrounding the lake, in particular the succulent *Selliera radicans* and the ground cover *Wilsonia rotundifolia.* The second class covers the range of native C_3 grassland communities adjacent to the lagoon, namely *Danthonia trenuior* and *Poa labillardierie* dominated areas. The third class covers the *Themeda triandra* remnant patches found on the western slopes of the site. The fourth class is representative of the scattered *Acacia* and *Bursaria* specimens found amongst the *Themeda* grassland, and the final class consists of exposed soils found within the lagoon.

2.1 Data collection

Data was collected using two different unmanned platforms mounted with different sensors. Spectral data was collected using a PhotonFocus 20 band hyperspectral snapshot camera, with

a spectral wavelength range from 600 to 875 nm and average Full Width Half Maximum (FWHM) of 6 nm. The camera houses a hyperspectral chip manufactured by IMEC with 25 band-pass filters mounted on top of the sensor's pixels in a 5x5 mosaic pattern. The 25 bands are captured simultaneously and the pixels are organised in a hypercube of 409 by 216 pixels, and resampled to 20 bands after spectral correction. The camera captured images at 4 frames per second (fps). The camera was mounted on a gimbal on a DJI S1000 multi-rotor UAS, and flown in a grid survey pattern at 80 m above ground level (AGL) with a flight line separation of 22 m providing 60% side overlap between flight strips. The flight track was recorded with a navigation-grade GPS receiver, and each hyperspectral image frame was geotagged based on GPS time. One hundred images were captured with the lens cap on the camera and averaged to collect a dark current image. Another 100 images were captured of a Spectralon panel directly before and after UAS flights to apply a vignetting lens correction and to allow for conversion of DN values to reflectance. A Python script was developed to process the raw camera data into hypercubes with reflectance values. The final band designations for the 20 output bands are shown in Table 5.1. The resulting images were exported to the GeoTiff format and imported into AgiSoft Photoscan (with their corresponding GPS coordinates). The SfM, dense matching, model generation, and orthophoto generation processing steps were performed in Photoscan based on band 14 (801 nm). Twenty-two photogrammetric ground control targets were randomly distributed across the study site and coordinates with a dual frequency geodetic-grade RTK GNSS receiver, resulting in an absolute accuracy of 2 cm. These GCPs were incorporated in the bundle adjustment step in Photoscan to facilitate accurate orthomosaic generation. A 347.8 m by 254.8 m hyperspectral orthomosaic of the full scene was produced for further analysis. Figure 5.1 shows the hyperspectral orthophoto, with a final spatial resolution of 14 cm, depicted as a false-colour RGB composite using bands 14, 5 and 1. Spectral values are shown for the subset areas of each class.

Additionally, an RGB camera was flown on a fixed-wing UAS at a height of 80 m. The data collected by this sensor was used to create a 15 cm spatial resolution digital surface model (DSM) in Agisoft Photoscan using the SfM workflow described earlier. From this DSM, slope was derived using the surface toolset in ArcGIS 10.3 (Environmental Systems Research Institute, 2014).

For validation purposes, two 100 m transects were established at the site during aerial data acquisition. The transects covered the *Wilsonia*, *Danthonia* and *Themeda* classes over an area in which the communities intergrade significantly. Transects were run east to west across the centre of the study area. Observations of plant communities were taken every meter along each transect. A polygon representing the observation area was then digitised in ArcGIS 10.3 for each observation site across the transect, and assigned the relevant class based on the field observations. Training points for the classification model were based on field observations acquired during the field campaign for Chapter 3, and a number of additional points were taken within the site based on ground reference points used for orthorectifying the UAS data. Each training point was buffered to 5 m in ArcGIS 10.3, as this was the size of the plots over which vegetation classes were identified in Chapter 3. Each buffered zone was converted to a series of points spaced 15 cm apart to match the spatial resolution of the orthophoto on which

classification was performed. Vegetation classes were assumed to be uniform within the entire 5 m zone, and care was taken to ensure that no points fell on transitional boundaries. In addition, a set of reference segments representing homogeneous class regions were digitised for the four vegetation classes to serve as both validation data and input for the segmentation assessment workflow described in chapter four. Figure 5.2 shows the distribution of the training points, validation transects and reference polygons within the study site.

2.2 Segmentation assessment

In order to identify the optimal segmentation scale for each vegetation community, the segmentation assessment method described in Chapter 4 was implemented. The 20 band hyperspectral orthomosaic, DSM, and slope datasets were combined and segmented 37 times using the multi-resolution segmentation algorithm in eCognition V 9 (Trimble Navigation, 2014). The initial segmentation was undertaken at a scale factor of 250, with a shape value of 0.9, and a compactness value of 0.1. As the plant communities found in the site are unmanaged, the distribution of community patches is not uniform, with many patches existing in elongated stretches around the edge of the lagoon. Therefore, it was determined that the creation of smooth rather than compact image segments would be of benefit in this case. Subsequent segmentations were performed by increasing the scale factor by 50 each time, until a maximum scale factor of 2000 was reached.

The segmentation validation process was undertaken in an identical manner to the case studies presented in Chapter Four. Optimal segmentation scales were predicted for the four vegetation classes. The soil class was excluded from this analysis as it exists in only a single large patch, meaning that only one large reference object could be created covering the entire class area. This would mean that the creation of independent training areas would not be possible, therefore, only thematic similarity was calculated for this class. Similarity values (S_{xc}), as defined by equations 4.1 and 4.2 were calculated between all image segments (X) for each thematic class c. Values for S_{xc} were calculated using the training points as the source of class definitions. Once S_{xc} had been calculated for the full range of X at each segmentation scale, the area of intersection between X and the digitised reference objects (R) was calculated. From there, values for $Con1_c$ and $Con2_c$. were derived. For a class, $Con2_c$ is defined as the average of S_{xc} value for all segments with a predicted class of maximum thematic agreement of class c that intersect a reference object also belonging to class c. $Con1_c$ is defined as the area of intersection between the aforementioned segments with correctly identified classes of maximum agreement and the parent reference object. Finally, the value of U representing the area of intersecting segment area located outside of the reference object boundaries was calculated for each segmentation scale (Weidner, 2008). Values for the three metrics were then combined to produce the final segmentation index value, S_q. The optimal segmentation scale was then determined for each class and for the scene as a whole by identifying the scale factor at which S_q reached its minimum value for each class respectively. The optimal segmentation scale factors predicted for each class are shown in Table 5.1.

Class-specific optimal segmentation scales, as well as predictions of overall optimal scale were used to produce predictions of class extent. Image segmentation was then performed a second

time on the 14 cm orthomosaic, DSM and slope datasets based the predicted optimal segmentation scales for each of the vegetation classes. The predicted extent for each class were imported into eCognition as thematic layers and then used to identify segments that had their highest thematic agreement equal to the given class. Segmentation was initially performed at the finest identified optimal segmentation scale, and then subsequent refinement was undertaken by re-segmenting the image at the second finest identified scale factor, but with segments identified correctly at the original scale factor being excluded. Segments correctly identified in the predicted extent layer for the class associated with the second finest segmentation scale were then also excluded, and third segmentation undertaken on the remaining segments at the third finest identified optimal scale factor. Finally, all correctly identified segments belonging to the class associated with the third segmentation scale were excluded, and the final segments re-segmented at the coarsest optimal segmentation scale. The output of this segmentation will be referred to as the multiple scale segmentation (MSS), due to the use of several scale factors in the final output.

Additionally, segmentation was performed using a single scale factor based on the overall optimal scale prediction as determined by S_q for the entire scene based on the mean index values for each class, and based on the total set of reference objects. These two methods of assessing scene-wide optimal segmentation scale predict a single scale factor for the entire scene. For the mean of all classes approach (MC), the index values derived for each class are averaged in order to give a final estimate of general performance at a given spatial scale. For the approach based on the total set of reference objects (TRO), index values are calculated for all segments intersecting segments with a correctly predicted class of maximum thematic similarity. This method provides an additional estimate of general class performance, which accounts for the overall variability in spatial scale of classes.

2.3 Random forest training and classification

Once the segmentation assessment process had been completed, a random forest model was trained for classification. Training and classification were performed on the 20 band orthomosaic, the DSM and the slope layers. As the number of input variables was equal to 22, the number of variables to try (mtry) was set equal to 4, as the established optimal parameter value is equal to $\sqrt{m}$ (Breiman, 2001; Naidoo *et al.*, 2012). Internal cross-validation accuracies were obtained for the model, in addition to variable importance measures. RF classification was performed on the MSS, MC and TRO segmentation results. Validation of the classification results was performed twice for each result; once using the reference objects used in the segmentation assessment, and a second time based on the field validated transects. The reference objects and transects were not merged into a single validation dataset due to the difference in the scale of analysis between the two datasets. The size of observation points along each transect was 45 cm^2, resulting in a significantly smaller observation area than that covered by the reference objects. Merging the transect observation areas into the larger reference segment area would therefore result in this information being lost due to the large discrepancy in the area of analysis between the two datasets. Validation was performed using the reference segments in order to provide a large-scale estimate of accuracy across the entire

scene. Validation using homogeneous reference segments also allows for evaluation of misclassification due to oversegmentation. The field transects were used as a secondary source of validation as they provide valuable data about the sensitivity of the classification results to transitional zones between communities. The high spatial frequency of observations along the transects allows for accurate determination of the exact point of change between vegetation types. As community intergrading has been identified in previous chapters as being a significant source of classification confusion, the decision was made to collect data capable of evaluating the sensitivity of the segmentation scale and classification approaches.

3 Results

3.1 Segmentation assessment

Table 5.2 shows the predicted optimal segmentation scales for each of the four vegetation classes as predicted by $Con1_c$, $Con2_C$, U, and S_q. Optimal scale factors for all classes are similar. Additionally, the values predicted for each class by $Con1_c$ are also similar, with the exception of the *Acacia* class which has a lower prediction. There is significant variation in the optimal scale factors predicted by $Con2_c$, with the *Danthonia* and *Acacia* class having much finer scale predictions than the *Wilsonia* and *Themeda* classes. This result is expected for both classes, as both classes consist of many isolated patches. The two scene-wide predictions for optimal segmentation scale are 400 for the TRO result and 1700 for the MC result

Figure 5.3 shows changes in index values for each class with increasing segmentation scale. In general, classes perform better in $Con1_c$ at larger segmentation scales. Changes in the values of index values between successive scale factors indicates changes in the thematic class predictions for intersecting segments, as segments with lower agreement values are excluded from subsequent calculations. Values for U increase as the segmentation scale becomes coarser, while for the *Danthonia/Poa* class, $Con2_c$ shows a decrease in metric values with segmentation scale, indicating an improvement in performance.

Figures 5.4, 5.5 and 5.6 show estimated class extents under the class-specific and scene-wide optimal segmentation scale predictions. The estimated extents are shown at the segmentation scales predicted by each of the three input metrics ($Con1_c$, $Con2_c$ and U) and the final combined metric (S_q). All segments within the scene that have their maximum thematic agreement to the given class are displayed. This provides a means of assessing the correctness of the scale predictions, and the potential spatial distribution of classes at each segmentation scale. Ideally, displayed segments will correspond to known distributions of the class, as this indicates that the segmentation scale is able to accurately represent the thematic and spatial attributes of the class

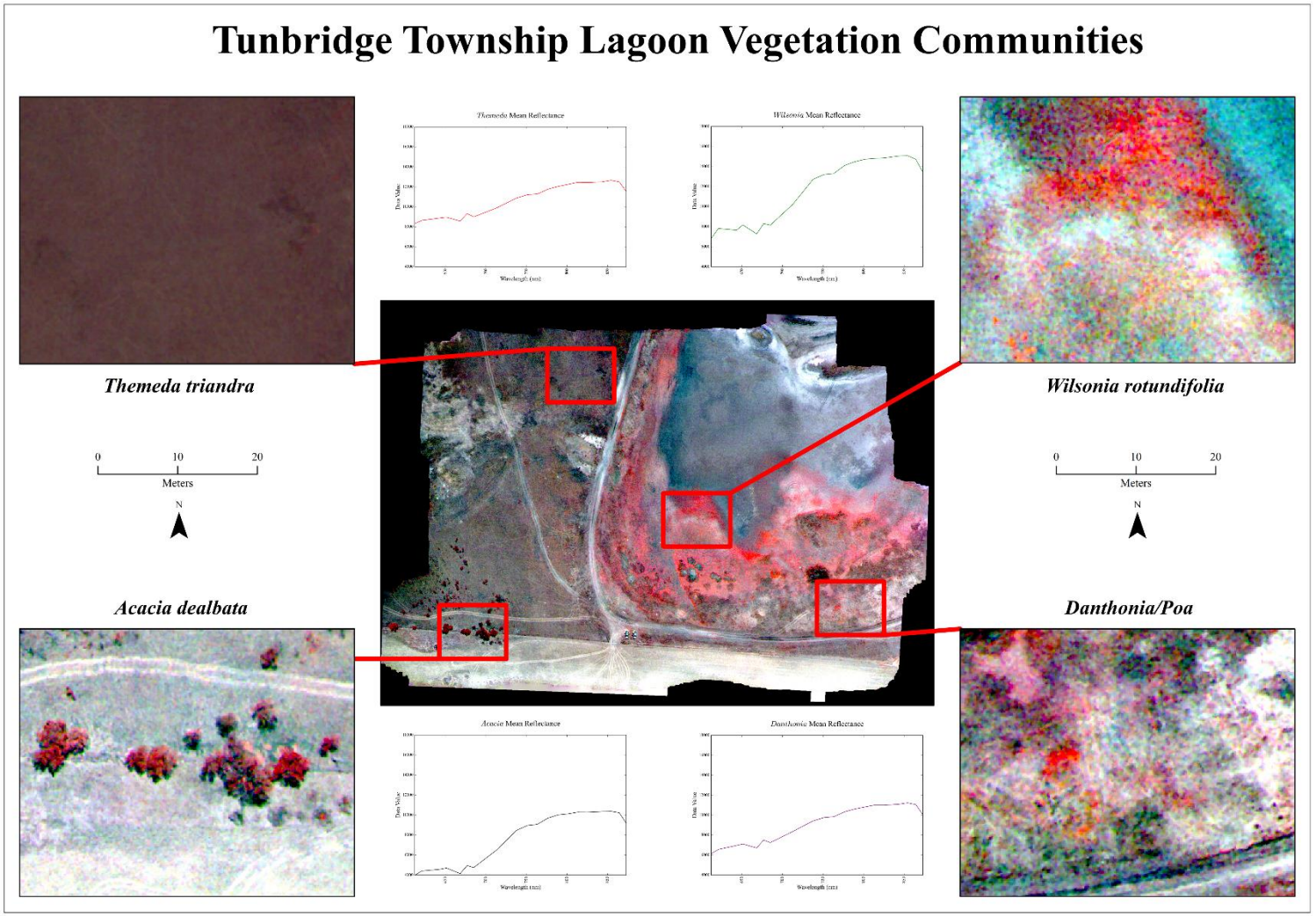

Figure 5.1: Vegetation communities found within the study site, as represented by the 14 cm orthomosaic. Central panel shows the overview of the study site, with image subsets showing general appearance of classes within the scene. Example spectral signatures are given for each class, based on the 20 spectral bands of the orthomosaic, as collected by the PhotonFocus hyperspectral sensor.

Table 3: Spectral band designations for the 20 band hypserspectral PhotonFocus orthomosaic

Band Number	Centre Wavelength (nm)
1	612.8
2	620.9
3	643.3
4	651.1
5	668.1
6	676.7
7	684.9
8	712.7
9	737.3
10	751.2
11	763.2
12	776.5
13	789.0
14	801.7
15	813.3
16	825.9
17	842.8
18	854.3
19	864.4
20	872.5

3.2 RF training and variable importance measures

Table 5.2 shows the confusion matrix and training accuracies obtained from the RF internal cross-validation. Class values are given as a pixel count, while accuracy is given as a percentage. The overall training accuracy was 97.44%. The obtained accuracies are high for all classes, with the *Themeda* and *Acacia* classes having slightly lower accuracies than the *Wilsonia* and *Danthonia/Poa* classes. There is very little confusion between classes, indicating good potential class separability within the dataset.

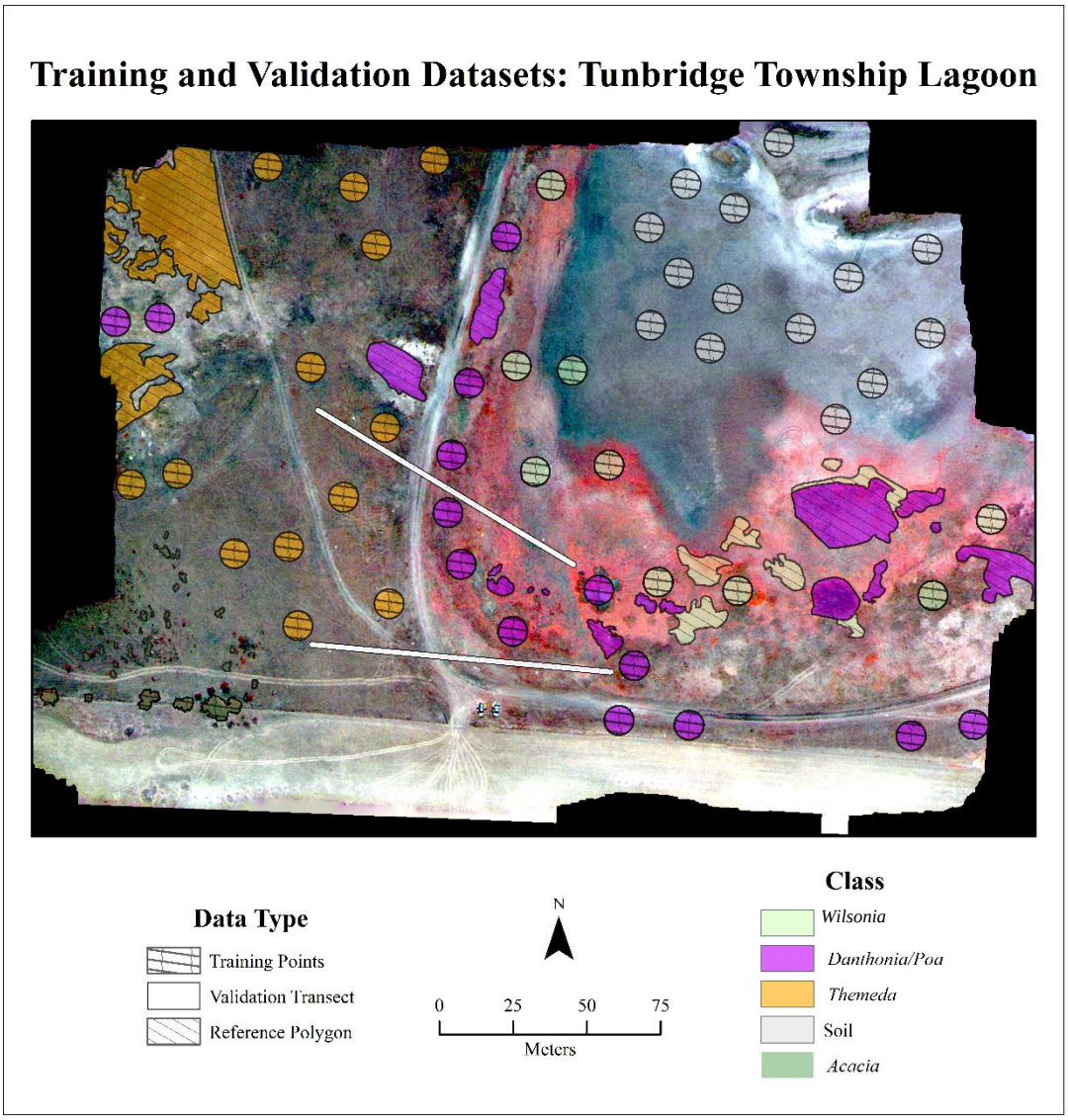

Figure 5.2: Distribution of training and validation points throughout the study site. The base image is a false colour composite of the 80 m PhotonFocus orthomosaic, where RGB=14,5,1

Table 5.2: Predicted optimal segmentation scale for all classes and the overall scene as determined by each of the segmentation assessment metric components. Ranges of scale factors are given for classes when one or more scale factors obtained identical values.

	$Con1_c$	$Con2_c$	U	S_q
Wilsonia	1900-1950	1700-1850	200	400
Danthonia	2000	300	200	350
Themeda	1450-1650	2000	200	200
Acacia	700-750	450	200	500
Mean	1550	1900-2500	200	400
All objects	1550	1700	200	1700

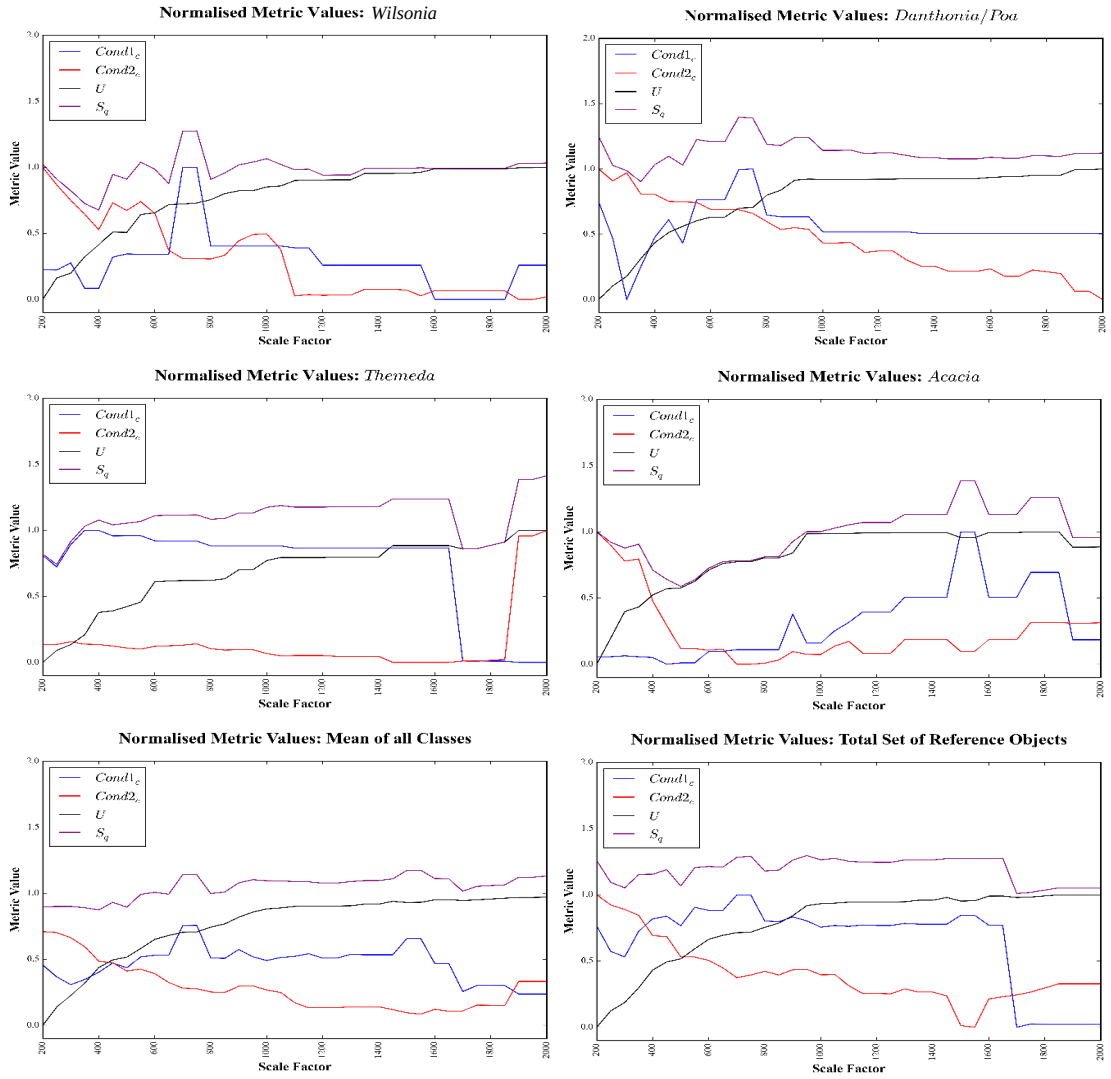

Figure 23: Plots of metric values derived from the segmentation assessment process. Values are plotted against segmentation scale factor for all classes. Additionally, the mean metric values over all input classes and the values obtained based on the total set of reference objects are provided.

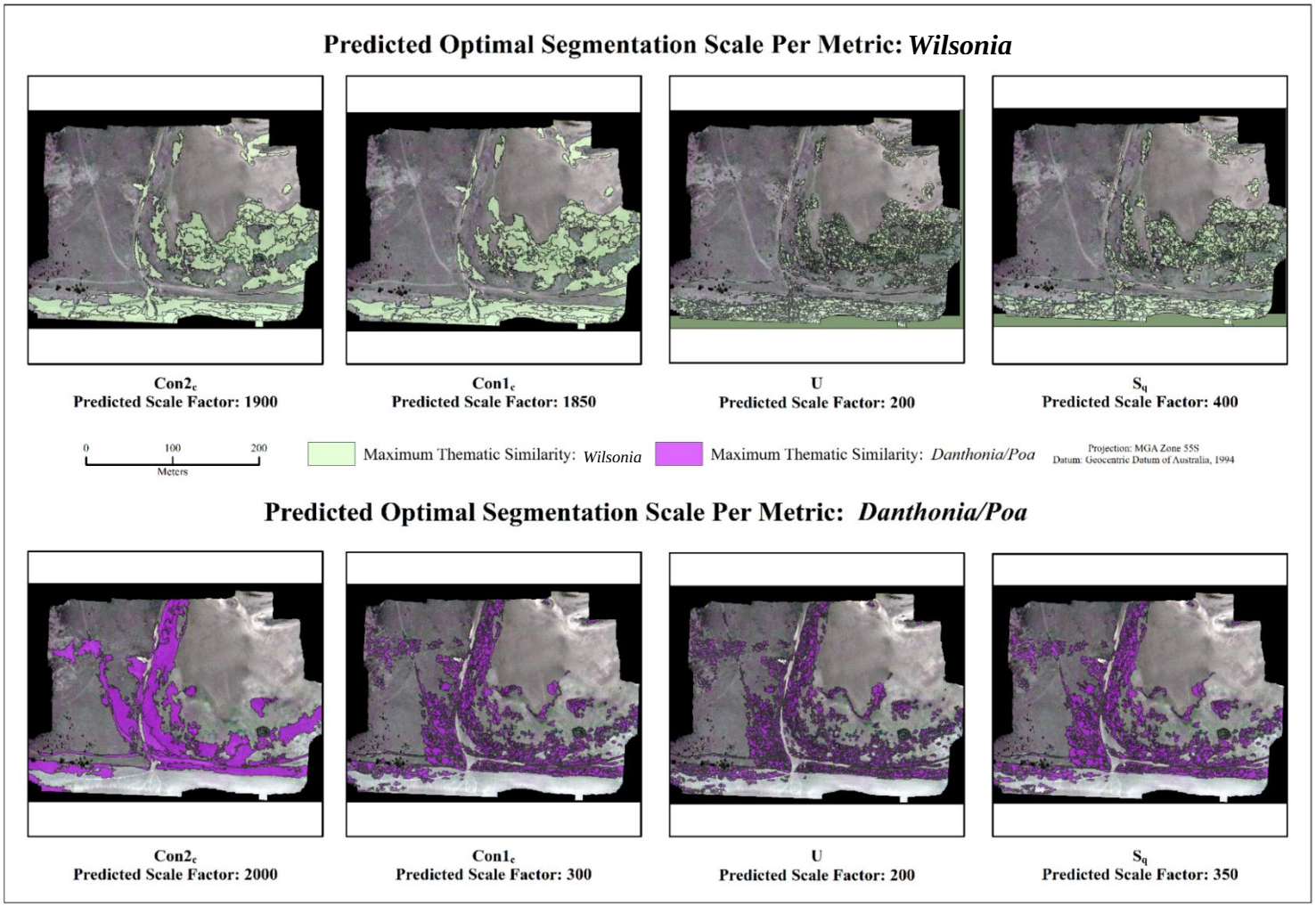

Figure 5.4: Predicted optimal segmentation scales for the *Wilsonia* and *Danthonia/Poa* classes as determined by all metrics. Shaded areas represent image segments that have their highest thematic similarity (S_{xc}) to the given class.

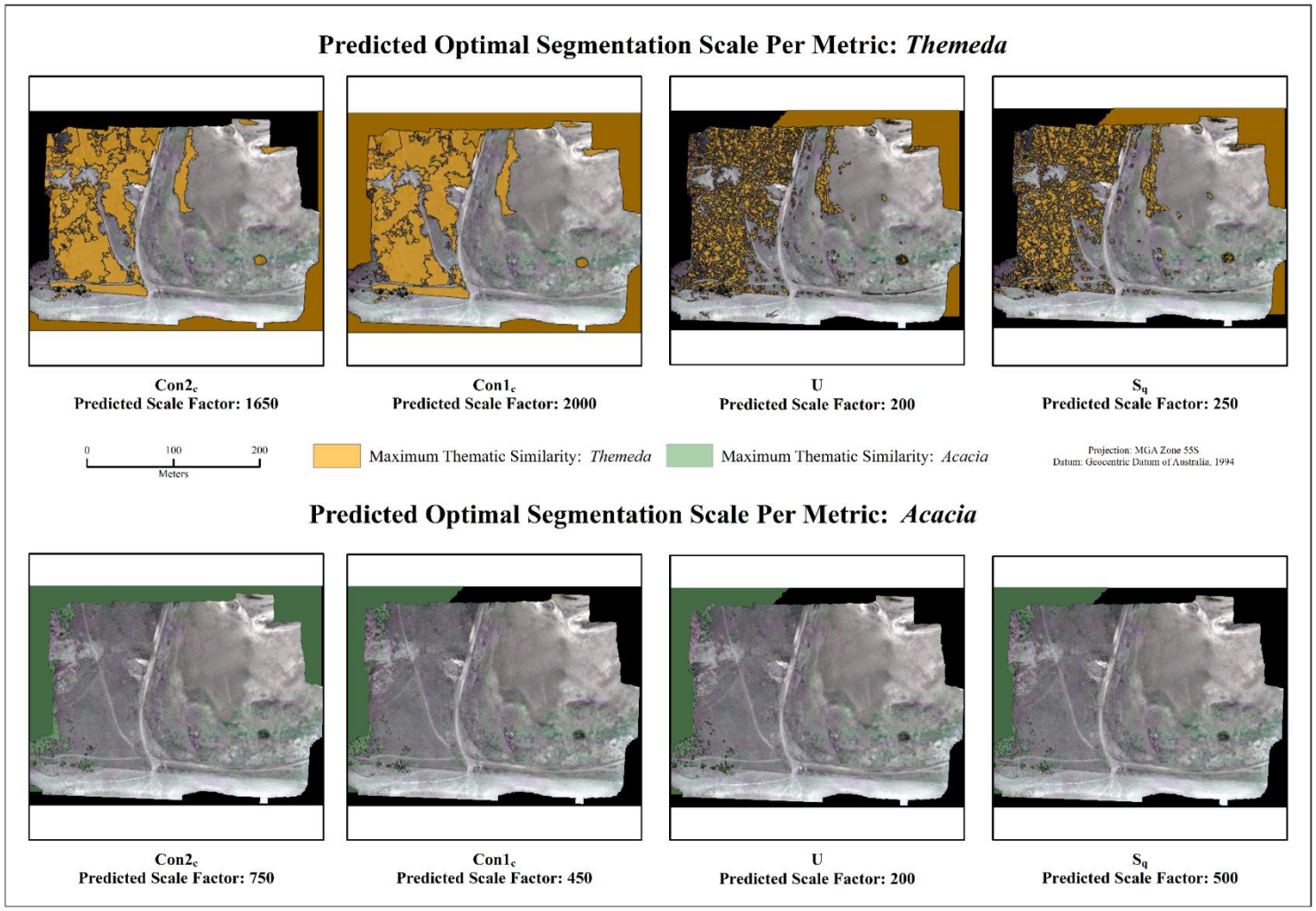

Figure 5.5: Predicted optimal segmentation scale factors for the *Themeda* and *Acacia* classes as determined by each component metric, and the final combined metric.

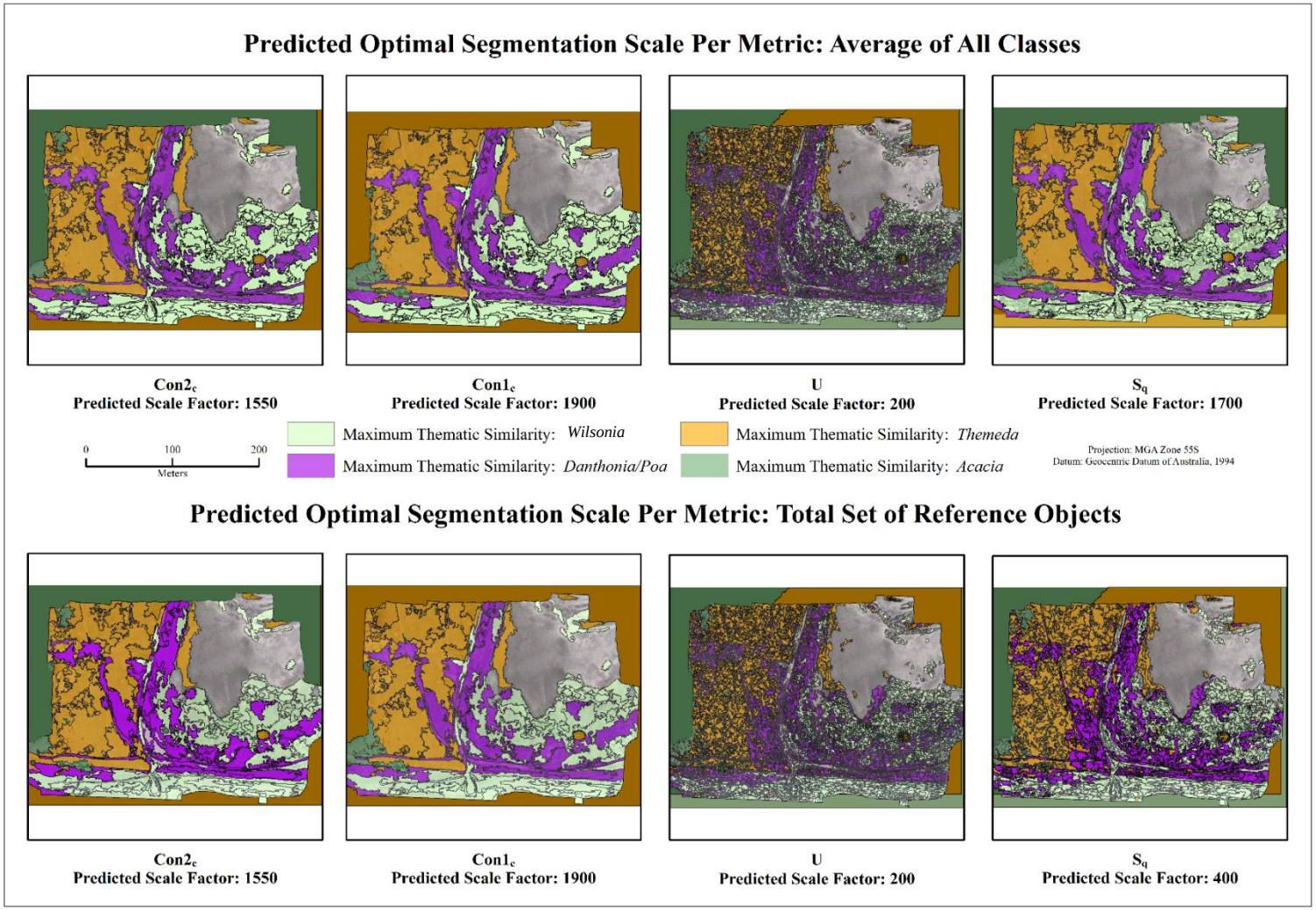

Figure 5.6: Scene-wide predictions of optimal segmentation scale derived based on the mean metric value of all classes (top row) at the given scale, and for the total set of reference objects (bottom row)

Figure 5.7 shows the variable importance measures obtained from the RF training model for each class. The DSM has a very high importance score relative to the other variables, and is identified as highly important for all classes. The most important spectral bands are bands two (620.9 nm), seven (684.9 nm), and eleven (763.2 nm). The *Danthonia* class has high importance values for these bands compared to the other classes. The *Themeda* class has the highest importance value for the DSM.

Table 5.3: RF training accuracy and confusion matrix for all classes. Confusion matrix values are given as a pixel count, while accuracy is reported as a percentage.

	Wilsonia	***Danthonia***	***Themeda***	**Soil**	***Acacia***	**Accuracy (%)**
Wilsonia	35,565	277	0	36	0	99.13
Danthonia	343	54,725	693	0	1	98.14
Themeda	0	3473	48,293	0	19	93.26
Soil	3	0	0	55,778	0	99.99
Acacia	0	123	345	0	7,740	94.29

3.3 RF classification accuracy

Figure 5.8 shows the RF classification results for the MSS segmentation based on class-specific predictions of S_q and the results from the two scene-wide segmentation assessment approaches; MC and TRO. Class extents are observed to be similar for all three segmentation scales, particularly for the MSS approach and MC. The result from the MC approach, at scale factor 1700 shows some discrepancies in class extent, particularly for the *Danthonia* and *Wilsonia* classes. Overall, class extents are well defined, and there is only a small degree of evident misclassification present in the *Acacia* class, which has several areas of over-predicted extent towards the north-west and south-west corners of the scene.

Table 5.4 presents the classification accuracy obtained for MSS segmentation. For all results evaluated against the reference objects, the soil class is included despite the fact that there are no validation areas associated with this class. Any segment identified as soil within the results is therefore misclassified, and included in the results for the sake of completeness. For MSS, mean classification accuracy is 70.6% when evaluated against the reference objects. User's and Producer's accuracies are reported for each class. The highest User's accuracy was achieved by the *Wilsonia* class, at 98.1%, while the highest Producer's accuracy was achieved in the *Themeda* class at 93.4%. The poorest User's and Producer's accuracies were both obtained in the *Danthonia* class, where both were equal to 37.3%. The poor overall accuracy of the class is the result of significant misclassification of *Danthonia* areas as *Wilsonia*.

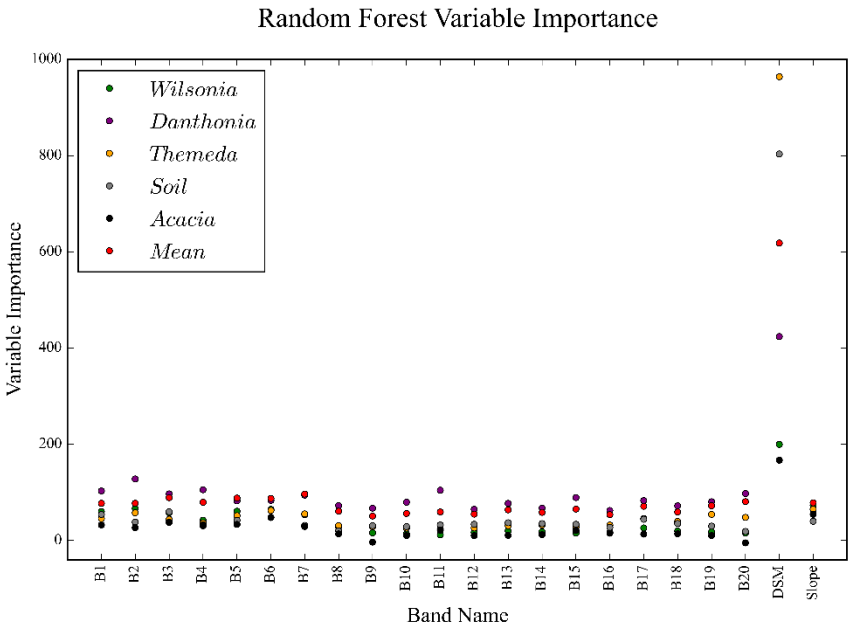

Figure 5.7: RF variable importance models obtained from the training model for each class

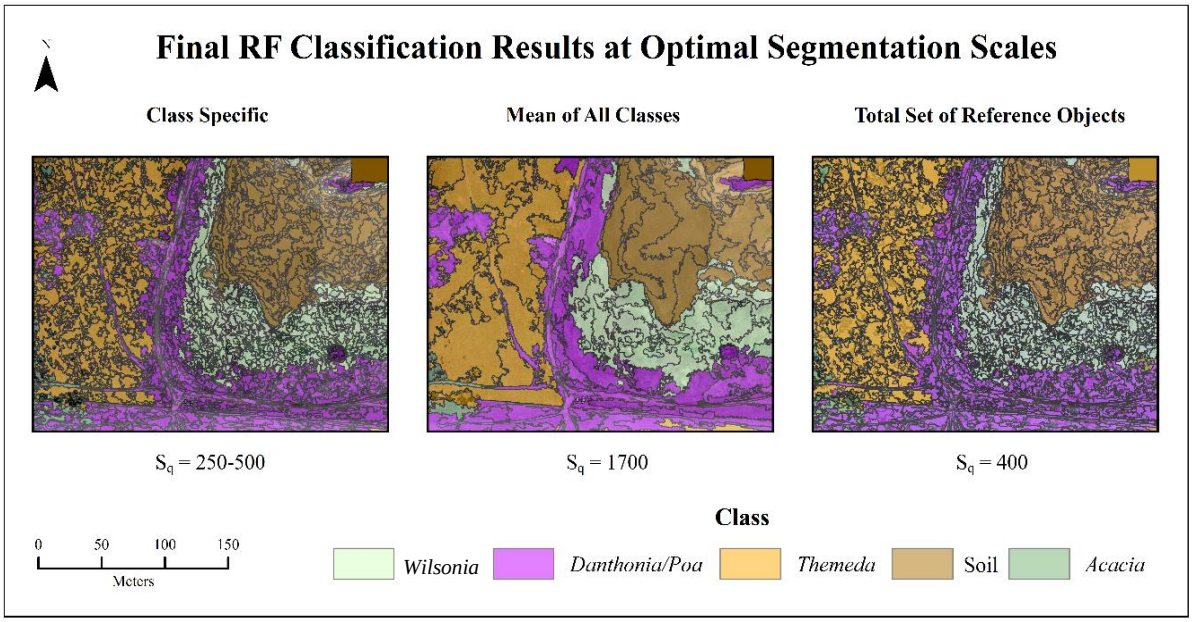

Figure 5.8: Final RF classification results based on the class-specific approach to segmentation derived from class values in S_q, and the results for the two single segmentation scale approaches; the mean S_q value across all classes, and the value of S_q obtained across the total set of reference objects regardless of class.

Table 5.4: Confusion matrix and final User's and Producer's accuracy for the MSS segmentation as determined using the reference segments. All class values are reported in m^2, while accuracies are reported as percentages.

	Wilsonia	***Danthonia***	***Themeda***	***Acacia***	**Soil**	**User's**	**Producer's**
Wilsonia	620.38	11.58	0	0	0.37	98.1	39.2
Danthonia	964.01	682.35	82.79	0	98	37.3	37.3
Themeda	0	23.79	1,699.59	96.13	0	93.4	93.4
Acacia	0	0.043	5.4	75.12	0	93.2	93.2

Table 5.5 reports the final classification accuracy achieved in the MC segmentation. The overall accuracy is equal to 72.2%. In this result, there is less confusion between the *Danthonia* and *Wilsonia* classes than in the MSS, although the overall User's accuracy for the *Danthonia* class is still poor at 46.6%. The Producer's accuracy for the class has improved significantly from the previous result. There are small decreases in classification accuracy in terms of User's accuracy for both the *Wilsonia* class and the *Themeda* class from the previous result, however Producer's accuracies for these classes are similar. The *Acacia* class has a significantly lower User's accuracy than that seen in the MSS result, at 57.4%, however the Producer's accuracy has increased to 100% from 93.2% in the previous results.

Table 5.5: Confusion matrix and final RF classification accuracy for the MC segmentation, evaluated using the reference segments.

	Wilsonia	***Danthonia***	***Themeda***	***Acacia***	**Soil**	**User's**	**Producer's**
Wilsonia	595.85	36.49	0	0	0	94.2	39.5
Danthonia	913.7	851.42	61.98	0	0	46.6	91.0
Themeda	0	163.78	1,655.1	0	0	91.0	94.5
Acacia	0	0	34.33	46.24	0	57.4	100

Table 5.6 gives the final RF confusion matrix and User's and Producer's accuracies for the segmentation scale factor of 400 result, which is the optimal segmentation scale factor predicted by with the total set of reference objects. The final overall accuracy for the result is 71.8%. The User's accuracies for the *Wilsonia, Themeda* and *Acacia* classes are very close to the results seen in the MSS result, at 98%, 92.4% and 92.6% respectively. Additionally, the Producer's accuracies for the *Wilsonia* and *Themeda* classes are also similar. The Producer's accuracy for the *Acacia* class is the lowest observed across the three results, at 41.9%, which is a large decrease from the high values seen previously. The Producer's accuracy for the *Danthonia* class has increased slightly over the result of the MC, to 98.4%,

Table 5.6: Confusion matrix, User's Accuracy, and Producer's Accuracy for all classes in the TRO segmentation.

	Wilsonia	***Danthonia***	***Themeda***	***Acacia***	**Soil**	**User's**	Producer's
Wilsonia	619.68	12.29	0	0	0.37	98.0	38.4
Danthonia	996.08	753.9	77.16	0	0	41.3	98.4
Themeda	0	0	1,680.8	103.46	0	92.4	95.3
Acacia	0.036	0	5.94	74.59	0	92.6	41.9

Table 5.7 shows the RF classification accuracy obtained for the MSS segmentation result based on the two validation transects. The overall accuracy for the classification is 80.8%. Compared to the results of the TRO, the User's and Producer's accuracies obtained for the *Danthonia* have improved significantly. The User's accuracy for the *Themeda* class is lower compared to previous results, at 65.6%. The classes most often confused with each other are consistent with the types of confusion seen in the previous results.

Table 5.7: Confusion matrix and final User's and Producer's accuracies for the MSS RF classification. All class confusion values are given as an area in m^2, while User's and Producer's accuracies are given as percentages.

	Wilsonia	***Danthonia***	***Themeda***	**User's**	Producer's
Wilsonia	5.55	0	0	100	71.3
Danthonia	2.23	12.36	0	84.7	77.1
Themeda	0	3.68	7.03	65.6	100

Table 5.8 shows the confusion matrix and classification accuracies obtained from the RF classification validation for the MC result using the validation transects. Mean accuracy is equal to 93%. Performance in this result is good, with high values obtained for both the User's and Producer's accuracies in all classes. Obtained values are higher than those seen in the multiple segmentation result, most notably in the User's accuracy for the *Themeda* class, and the Producer's accuracy for the *Danthonia* class.

Table 5.8: RF classification accuracy for the MC result evaluated against the validation transects

	Wilsonia	***Danthonia***	***Themeda***	**User's**	Producer's
Wilsonia	5.55	0	0	100	81.5
Danthonia	1.26	13.05	0.27	89.4	95.5
Themeda	0	0.62	10.1	94.2	97.4

Table 5.9 reports the validation results based on the transects for TRO classification. A mean User's accuracy of 85.1% was achieved. The type of confusion is similar to that observed in the other two evaluations, as well as for the validations undertaken using the reference objects. Primarily, confusion occurs between the *Wilsonia* and *Danthonia* classes. Overall, class-specific accuracies are similar to those described for the multiple segmentation scale result

Table 5.9: Class-specific classification accuracies for TRO result as evaluated using the transects

	Wilsonia	***Danthonia***	***Themeda***	**User's**	**Producer's**
Wilsonia	5.55	0	0	100	73.2
Danthonia	2.03	12.55	0	86.01	83.1
Themeda	0	2.55	8.17	76.2	100

4 Discussion

4.1 Segmentation Assessment

The extent predictions for the vegetation classes shown in figures 5.4 to 5.6 show a high degree of similarity. Within classes, the extent of vegetation identified based on the optimal scale predictions of each metric are very similar even when the scale factors identified differ significantly. For example, in figure 5.3, the scale factor predicted by $Con2_c$ for the *Wilsonia* class is equal to 1900, while the scale factor predicted by S_q is equal to 400. The number and size of segments visibly differs between the two scale factors; however, the general class extent is similar. This can be observed across the range of classes, as well as for the two scene-wide assessments shown in figure 5.5.

Generally, there is a tendency for the optimal scale factors predictions based on $Con1_c$ and $Con2_c$ to be highly generalised, as evident in the predicted class extents shown in Figures 5.5 to 5.7. This tendency was observed previously in the results of Chapter Four. The reason for this generalisation is likely to be related to increased segment size, in that larger segments incorporate a wider range of pixel values. A wider range of values means that summary statistics derived for individual segments will be less susceptible to outliers and other potential error sources. This in turn means that segments are better able to approximate class distributions, and therefore have lower overall levels of disagreement when the degree of similarity between segments and classes is assessed. The values predicted by U are identical for all classes, with optimal predicted scale consistently being the lowest of the tested scale factors. This result was observed previously in Chapter Four. As smaller segmentation scales produce smaller objects, the number of intersecting segments will be higher, but the area of individual segments smaller than at larger segmentation scales. This means that there is less opportunity for undersegmentation, as single scene objects are likely to be oversegmented and represented by multiple segments. These multiple segments are better able to delineate the boundary of the reference object than larger, more generalised segments. Although the occurrence of oversegmentation is not ideal, and can still result in misclassification, it is generally accepted as less of a concern than undersegmentation (Blaschke *et al.*, 2014). As the three individual metrics are combined to form the final predicted optimal segmentation scale, the tendency of U to predict overly fine segmentation scales acts as a counter-balance to the previously described over-prediction seen in the results for $Con1_c$ and $Con2_c$.

The class extent predictions highlight several issues in regards to the accuracy and representativeness of the training datasets. The *Acacia* class has significant over prediction of

extent in the upper north-western and south-western corners of the image. These areas are darker in appearance than the surrounding scene, and can appear similar to the typically dark foliage of the *Acacia* trees. The cause of this darkness is likely to be image vignetting, or edge effects arising during orthomosaic generation. As these areas form the extreme edges of the scene, the number of collected images is lower than for more centralised areas, resulting in less accurate pixel values. There are also significant areas of misidentification in areas of the *Themeda* class around the south-west foot of the hill. This area was burned in the summer of 2014, and is still regenerating, which has resulted in the area being identified as *Danthonia/Poa* instead of *Themeda*.

4.2 RF Training results and variable importance measures

The accuracies obtained for the RF training model are much higher than the accuracies obtained for any of the final classification results. The presence of significant discrepancies between validation and training accuracies can indicate potential bias in the sampling regime, or unrepresentative training datasets. As the number of input points was high (~250,000), it was decided that a single RF model was to be derived. The high spatial resolution of the dataset (14 cm) means that very fine-scale variations in species composition can potentially be detected. As validation points were derived over a 5 m plot area to emulate the conditions used in chapter three, there is potential to incorporate multiple thematic classes within a single plot.

The variable importance measures obtained show a very high importance for the two topographic variables, which was observed in the variable importance measures derived for the Landsat ETM+ and WorldView-2 RF models used in Chapter 2. Spectral variables have significantly lower importance values for each class, and all classes have their highest importance level recorded for the DSM. The high importance value assigned to the DSM in the *Themeda* class is due to the class occurring exclusively on the hilly area within the study site, whereas as the *Danthonia* and *Wilsonia* classes occur on the flat saltpan surrounding the lagoon. This difference is of key importance due to similarities in canopy structure between the *Themeda* and *Danthonia* classes, which were identified in Chapter 3 as being a factor contributing to misclassification between the two classes.

Classes exhibit different importance levels across the range of spectral input bands, with the *Danthonia* class having the highest overall spectral importance values. Key bands of importance for this class include band two (620.9 nm), 11 (763.2 nm) and 14 (801.7 nm). All other vegetation classes have comparatively low spectral importance values. The high values for spectral bands in the *Danthonia* result is likely due to similarities between the physical distribution of the *Danthonia* and *Wilsonia* classes, as both only occur on the same flat region of saltpan. As they two classes intergrade significantly, and cannot be differentiated based on topography, spectral bands are the only available source for differentiation with the RF model. This is likely to contribute to the poor performance of the class overall, as previous results indicate great difficulty discerning between the *Danthonia* and *Wilsonia* classes due to their similar photosynthetic pathway and phenological staging.

4.3 RF classification and validation

The areas of misclassification within the results are nearly identical to the areas misidentified in the predicted class extents derived from the segmentation assessment process. The final community extents obtained from the three different segmentation scales are very similar, however, there is some disagreement between the results for MC and the other two results. The disagreements in extent is primarily associated with the *Danthonia* class, particularly surrounding the southern boundary of the lagoon. The results obtained for the MSS and TRO results delineate this area well, however this is not the case for the MC results, in which the area is significantly overgeneralised. Additionally, there is still significant overprediction of the *Acacia* class in the north-westerly and south-westerly corners of the scene in all results.

The primary source of inaccuracy in the classification results across both validation tests is confusion between the *Danthonia* and *Wilsonia* classes. The confusion in this case is only in one direction, in that a large proportion of the *Danthonia* class is erroneously classified as *Wilsonia*, while there is very little misclassification of *Wilsonia* as *Danthonia*. As the two communities intergrade extensively, the establishment of discrete reference objects for segmentation validation was very difficult. Inaccuracy in the creation of these reference objects is likely to be the case of the poor overall accuracy obtained for the *Danthonia* class, however, the results of Chapter 3 indicated that there is significant difficulty in differentiating the two classes based solely on spectral properties. As the two classes additionally occur in the same area, primarily on low-lying saltpan, the DSM and slope variables are not likely to increase separability between the two classes. The *Danthonia* class exhibits significantly different final classification accuracies between the validation performed using the reference objects and the validation based on the image transects. The significant increase in classification accuracy in MS and TRO results based on the image transects over the results for the reference segments indicate the need for highly detailed fine-spatial scale validation datasets in order to accurately assess community intergradation. Additionally, the segmentation scale factors predicted as optimal for the *Wilsonia* and *Danthonia* classes are very similar, indicating similar spatial scales of occurrence.

Based on the validation using the reference segments, the most accurate results were obtained for TRO segmentation. The MSS segmentation obtained similar, but slightly lower final classification accuracies. The scale factors used to generate the MSS inputs are all close to a value of 400. The classification accuracies for the MC segmentation are also similar to the accuracies of the MSS and TRO segmentations. When the MC segmentation is validated against the transects, it achieved the highest overall classification accuracies of any of the results. However, the degree of overgeneralisation present in this result is concerning, and therefore it can be determined that the scene-wide optimal segmentation scale is best predicted using the TRO approach, rather than MC approach.

5 Conclusions

This chapter presents the results of an RF classification approach for identifying lowland native grassland communities in the Tasmanian Midlands using high spatial and spectral resolution

UAS imagery. Class-specific optimal segmentation scales were predicted and subsequently tested on the data. The results of the segmentation assessment indicated very similar optimal segmentation scales for classes, with scale factors ranging between 250 and 500. This gives a range in mean object size between 29.17 m^2 at scale factor 250, and 93.28 m^2 at scale factor 500. Scene-wide segmentation scales were also predicted using the two approaches identified in Chapter 4. Overall, the scene-wide prediction of optimal segmentation scale derived for TRO produced a very similar prediction of class extent to MSS result generated using the class-specific optimal scale factors. The behaviour of the various component metrics used to calculate the final optimal segmentation scale S_q is in keeping with the behaviour seen in the case studies presented in Chapter Four. There is a tendency for the scale predictions identified by U give an oversegmented result, as was also seen in Chapter Four. The final predictions based on $S_{q,}$ however, are generally very good, and indicate that the segmentation assessment method is capable of performing well in environments with a high degree of community intergradation.

Classification accuracies obtained for the three input segmentations are very similar for both validation results. Overall, the optimal scale factor predicted using TRO scale factor of 400 performed the best across the reference object validation with a mean accuracy of 72.4%, while the scale factor prediction based on MC performed best based on the transect validation with a mean accuracy of 93%. The results indicate that there is potential inaccuracy in the training points, particularly in regards to the *Danthonia* and *Wilsonia* classes. The higher classification accuracies obtained for the transect validations indicate that accurate assessment of community gradients requires the collection of high spatial frequency field observations over a large area. The small extent covered by the two transects means that assessment of communities in this manner is limited for this result. Future studies could benefit from the creation of multiple transects with closely spaced observations to aid in a more robust assessment of classification results.

6. Thesis Context

This chapter presents an approach to lowland native grassland mapping that combines the findings of previous chapters. Accuracies achieved in this chapter indicate that there is a need for high spatial resolution datasets in all stages of analysis. This means that training and validation data need to be collected using field-based approaches in order to provide accurate assessment of classification results. The findings of this chapter reiterate previous findings, in that there is significant difficulty differentiating C_3 dominated lowland native grassland species from other native C_3 native vegetation types. The importance of topographic variables within the classification workflow is also reiterated in these findings. This chapter also provides a case-study showing the potential applications of snapshot hyperspectral sensors for ecological research applications. This chapter has shown that these new sensors can function well, and provide high quality datasets for analysis.

Chapter 6

Conclusions

The aim of this thesis was to develop remote sensing methods for identification and mapping of endangered lowland native grassland communities in the Tasmanian Midlands region. The key goal was to identify methods capable of community identification that could contribute to frequently updated maps that can be employed for long-term monitoring of both status and extent. As current mapping methods are based primarily on manual digitisation of aerial photography, there was a need to assess the applicability and utility of various remotely sensed data sources in order to determine the optimal approaches for community mapping. This raised a number of questions regarding the selection of datasets. Classification tests were performed on different sources of remotely sensed data in order to determine the applicability of different spatial and spectral sensor resolutions for lowland native grassland community mapping. Additionally, the selection of an optimal scale of analysis has been frequently highlighted as an issue for accurate mapping of lowland native grassland communities. Here, a method for the identification of optimal image segmentation scale is proposed in order to mitigate this issue. The methods investigated, developed, and employed in this thesis have demonstrated that remote sensing is a viable option for lowland native grassland community identification and mapping provided that the spatial and spectral resolution is appropriate. This thesis has identified several key factors that contribute to success in community mapping, which can be used to produce accurate maps of lowland native grassland extents.

6.1 Multispectral approaches

Objective 1

To determine the suitability of moderate to high resolution multispectral satellite imagery for identification and mapping of lowland native grassland communities. Additionally, the ability of pre-existing coarse resolution vegetation maps to act as reference datasets was assessed. Multispectral satellite datasets are easily acquired and frequently captured, and many vegetation maps have been created for the region through manual digitisation of aerial photography. The application of these datasets within a new mapping framework is potentially capable of providing broad-scale analysis of communities that can be repeated with greater frequency than current manual and field-based methods. By utilising existing vegetation maps as reference data sources, the need for expensive field work can potentially be reduced.

Chapter 2 aimed to determine whether lowland grassland community types could be successfully classified using moderate and high spatial resolution multispectral satellite datasets. Additionally, the utility of existing data sources to act as classification and validation data for analysis was tested. Object-based image analysis was undertaken using a Landsat ETM+ scene and a WorldView-2 scene acquired over a property near Campbell Town in the Tasmanian Midlands. Classification was undertaken using a Random Forest approach. Training and validation was performed using a randomly generated set of points sampled from a coarse-resolution community map. The input reference dataset was created by the Tasmanian

Land Conservancy, and is based on digitisation of community extents based on field observations undertaken by a field officer. A *k*-fold cross-validation approach to training and validation was undertaken, with a total of 50 random subsets of points used for classification. The Landsat ETM+ and WorldView-2 images were then classified in conjunction with a digital elevation model and derived topographic variables. The remaining points not used in training were used to validate the corresponding models in order to ensure independence between the training and validation process. Analysis of Variance (ANOVA) was undertaken to determine if vegetation classes had significant differences in classification accuracy between the two datasets. Finally, the classification results were compared to the current vegetation map of Tasmania, known as TASVEG, and accuracies based on the range of validation points compared.

The classification accuracies achieved indicate that delineation between *Poa labillardierie* and *Themeda triandra* communities is possible using multispectral datasets, with mean classification accuracies found to be 71.6% and 71.77% respectively for the Landsat ETM+ results, and 73.19% and 76.79% for the WorldView-2 results. Differentiation of the third lowland native grassland community, known as the grassland complex, was comparatively poor, averaging 55.4% in the Landsat ETM+ results and 55% in the WorldView-2 results. The source of this marginal result is a high level of confusion with the dry woodland class. The ANOVA results indicate that there is significant improvement in classification accuracy for the *Themeda* and dry woodland classes based on the WorldView-2 dataset over the Landsat ETM+ dataset. Additionally, all classes achieve statistically significant improvements in classification accuracy over the TASVEG dataset.

Overall, the results of this study indicate that multispectral approaches to lowland native grassland community classification are feasible. Additionally, pre-existing datasets can be an appropriate and convenient source of input data for such approaches. The results of this study provide the first remotely sensed maps of lowland native grassland distribution for the region, and serve as an important step towards the development of a larger comprehensive monitoring scheme. The findings of this study present an important case-study within the field of grassland mapping, presenting the feasibility of thematic community mapping using a moderate resolution multispectral dataset. Additionally, the *k*-fold approach for cross-validation of the classification is a novel method within the field of vegetation mapping. The ability to statistically validate improvements in classification accuracy between input datasets is an important finding that provides extra resilience to the interpretation of classification results. This approach can be used across many potential applications, and can aid in both reducing sampling bias within classification results, and by providing increased certainty in the interpretation of the results. Both of these findings are important advancements for the field of image classification.

6.2 Spectral characterisation of lowland native grassland communities

Objective 2

To identify the spectral properties of remnant communities in order to determine the best approach for spectral differentiation. The utility of narrowband and broadband approaches needs to be tested in order to determine the required spectral resolution for accurate community identification. Knowledge about the optimal spectral resolution for separating dominant grassland species will aid in the selection of future data sources.

The goal of chapter three was to determine whether high spectral resolution datasets could provide improved classification results for lowland native grassland communities. Analysis was undertaken using high spectral resolution field data collected at the Tunbridge Township Lagoon using a handheld spectroradiometer. The original 501 spectral bands between 450 nm and 950 nm were run through a feature reduction method to identify and remove highly correlated and redundant bands. Additionally, the spectra were resampled to emulate the spectral resolution of the Landsat OLI and WorldView-2 multispectral sensors. Classification was undertaken using random forest models, and a *k*-fold cross-validation approach to training and validation. Spectra were classified using a four-class and three-class configuration to determine whether classification accuracy could be improved through generalization of C_3 vegetation types into a single class. Classification was performed for each configuration based on the full high spectral resolution dataset of 3 nm FWHM bands, a reduced high spectral resolution model derived using the feature reduction method, as well as the simulated Landsat-8 and WorldView-2 bands. ANOVA was then run to determine whether classification accuracies differed significantly based on the input dataset and between class configurations.

Variable importance measures derived from the range of RF training models showed strong selection of variables in spectral bands associated with pigment levels, water content, and photosynthetic levels. In particular, bands in the region of 550 nm associated with Anthocyanin content were identified for the *Themeda* class. Classification accuracies achieved were good across the majority of classes, although in the four-class configuration there was significant confusion between the *Danthonia* class (which is the dominant grass species group in the lowland native grassland complex class defined by TASVEG) and the groundcover *Wilsonia rotundifolia*. This result is similar to those seen in Chapter Two, in which the grassland complex was significantly confused with other vegetation classes with the same photosynthetic pathway.

The ANOVA results in Chapter 3 showed statistically significant improvement in classification accuracy for the merged C_3 class (composed of the original *Danthonia* and *Wilsonia* classes). It was found that the *Themeda* class had significantly improved classification accuracy in the reduced high spectral resolution dataset when a full four-class configuration was applied. Additionally, the *Danthonia* class was found to have significantly lower classification accuracy based on the Landsat-8 and WorldView-2 bands than for the full 501 band high spectral resolution dataset. For all other cases however, it was determined that the use of high spectral resolution datasets has a limited effect on overall classification accuracy. These results indicate that the use of high spectral resolution datasets does not provide statistically significant

improvements in classification accuracy over broadband multispectral approaches for the majority of lowland native grassland vegetation studied here. However, it should be noted that these results represent a small-scale study of grassland communities, and analysed spectra did not cover the full range of the electromagnetic spectrum. These results serve only as preliminary findings, and warrant further investigation into the applicability of high spectral resolution datasets. An additional key finding of this chapter is the identification of key spectral bands in which lowland native grassland communities can best be separated from each other and accurately identified.

The findings of this chapter are in keeping with those of the previous classification results presented in chapter 2. The spectral regions identified as important for community discrimination are similar to those found by the variable importance measures obtained by the RF model training stage of chapter 2. The identification of similar key spectral regions across both datasets is an important finding, as the two studies were undertaken at different locations within the Midlands region, and represent two different land management schemes. These similarities indicate that large scale mapping based on the identified spectral regions is potentially feasible, although larger-scale testing is still required.

6.3 Segmentation assessment

Objective 3

To develop a method for predicting optimal segmentation scale for multiple thematic classes that considers both geometric and thematic attributes of image segments and reference objects. The limited success of previous mapping attempts for lowland native grasslands can be attributed primarily to inappropriate scales of analysis; therefore, there is a need to determine the optimal image segmentation scale to match the spatial scale of remnant grassland community patches.

Chapter four presents a novel method for predicting optimal segmentation scale based on thematic and spatial attributes. The method utilises three metrics to assess various segmentation properties, which were then combined to provide a final prediction of optimal segmentation scale for individual classes. The degree of thematic similarity between individual image segments and each thematic class to be used in the final classification was assessed. For each segment, the class to which thematic agreement is highest was determined based on the calculated similarity values. Image segments were then intersected with reference objects, and the class of highest thematic agreement was compared to the class of the intersecting reference object. This comparison forms the basis for the development of two new metrics used to assess the thematic accuracy of a given segmentation. The first metric, $Con1_c$, was assessed for each class in turn. For all reference segments composing a given class, the area of intersection between reference objects and image segments with their highest thematic similarity to the same class was tabulated. $Con2_c$, was then tabulated based on the same identified segments. $Con2_c$ is the mean similarity metric value for all correctly identified intersecting segments for a class. The third metric, *U*, measures the spatial agreement between intersecting image segments and reference objects, and is equal to the percentage of total intersecting segment

area located outside the boundaries of reference objects. This area was calculated for each class. The three assessment metrics were then combined to provide a final estimation of optimal segmentation scale for each class. Scene-wide optimal scale was also calculated using two approaches. The first approach simply averaged the metric values for each class at a given segmentation scale, while the second approach calculated each metric value for the total set of reference objects. In this case, values were calculated based on any correctly identified intersecting segment regardless of the class to which it has its highest thematic similarity.

The segmentation assessment was applied to two different datasets in order to assess performance. In the first study, the method was applied to a 30 cm aerial photograph of an urban environment. Optimal scale factors were predicted for four thematic classes occurring at varying spatial scales. In the second case study, the method was applied to a 1.7 cm spatial resolution UAS image acquired at Tunbridge Township Lagoon. In this example, optimal scale factors were predicted for two classes; *Acacia dealbata* and *Lolium perenne.* For both case studies, predictions of optimal scale factors were good. Vegetation community boundaries were clearly delineated based on the class predictions, and class extents were similar to known class distributions. Overall, the method proved to be a highly effective in predicting class-specific optimal segmentation scale across varying environments.

The prediction of class-specific segmentation scale is a novel-concept within the field of object-based image analysis, and presents a significant advancement in the field. The ability to empirically justify the choice of segmentation scale used across an image has been a goal within the field for several years. The need for multiple segmentation scales for various thematic classes has been identified as being of key importance, and the results of this chapter show that there is potential to provide reliable estimates of optimal scale factors for individual thematic classes. Additionally, this chapter has shown that this approach is consistent across both naturalised and developed landscapes, which is a consideration that many segmentation assessment approaches have not addressed.

6.4 High spectral and high spatial resolution community mapping

Objective 4

To test the feasibility of unmanned aircraft systems (UAS) as a platform for collecting data at the optimal spatial and spectral resolutions identified by previous objectives. This objective aims to test the optimal data characteristics identified under previous objectives, and bring them together within a single case study.

Chapter 5 presents the results of a high spectral and high spatial resolution classification approach for lowland native grassland mapping. A 14 cm orthophoto was created using data collected with a PhotonFocus 20 band hyperspectral snapshot sensor mounted on a UAS platform. In addition, a Digital Surface Model (DSM) and slope model were derived using structure-from-motion. A total of four vegetation classes, *Wilsonia, Danthonia/Poa, Themeda* and *Acacia* were identified, as well as a soil class. Segmentation assessment was performed using the method described in Chapter 4 to identify optimal segmentation scale factors for each

of the four vegetation classes, as well as the scene as a whole based on segmentations performed using the multi-resolution segmentation algorithm in eCognition. The predicted scale extents for each class were consistent across the three segmentation assessment metrics, and conformed to expected community distributions. Once scale factors had been derived for each class, image segmentation was performed again using the optimal scale factors predicted for each class to create a multiple scale segmentation result. Predicted optimal scale factors ranged between 250 and 500 across the range of classes, while scene specific estimates of 1700 for the mean class performance and 400 for the total set of reference objects were also obtained. RF classification was performed on the multiple scale segmentation, the scale factor 1700, and the scale factor 400 outputs. Variable importance measures obtained from the RF training results were similar to those obtained in Chapters 2 and 3. Variables associated with photosynthetic activity levels as well as topography were identified as being the most important. The results also indicated poor spectral separability between the *Danthonia* and *Wilsonia* ground cover class.

Evaluation of classification results was performed using the reference objects digitised for each class for the segmentation assessment. Additionally, data were validated against two 100 m transects for which observations were recorded every metre. The classification accuracies obtained were similar in range to those obtained in previous chapters. The *Danthonia* class exhibited poor classification accuracy, with User's Accuracies between 37 and 47%, when evaluated against the reference objects. Confusion between the *Danthonia* class and the *Wilsonia* class was again the reason for such poor classification results. Overall classification accuracy values for the three input segmentations were similar, ranging between 70% and 72%. Classification accuracies for all classes when assessed against the validation transects improved significantly. The best results were achieved in the scale factor 1700 result, where class accuracies varied between 89% for the *Danthonia* class, and 100% for *Wilsonia*. The multiple segmentation scale result had the poorest results when validated against the transects, with a mean accuracy of 80.8%.

The results of this study indicate the need for spatially detailed validation data capable of accurately identifying community transition zones. The segmentation assessment method provided good overall predictions of class extent, however, the multiple segmentation scale result did not perform as well as anticipated in the final classification results. The high degree of generalisation observed in the training data could be a potential reason for this, as well as inaccuracies in the digitisation of reference polygons. Overall, the accuracies obtained are in keeping with those obtained in previous results, indicating that UAS system are a feasible form of data collection for lowland native grassland mapping. The combined approach of high spatial and high spectral resolution analysis has potential to provide accurate community classifications, but only if training and validation data are captured at an appropriately fine spatial resolution.

6.5 Contributions to knowledge

The contributions to knowledge of this thesis fall into two major categories; grassland remote sensing and object-based image analysis approaches. For grassland remote sensing, the major contributions of this thesis are:

- Improved understanding of requirements for spatial and spectral resolutions needed to differentiate various lowland native grassland communities.

- Confirmation that moderate spatial resolution multispectral approaches to community mapping are feasible, which allows for less time between successive community maps. This is a significant finding within the larger context of Tasmanian vegetation mapping, as it allows for updating current methods and improvements in map accuracy.

- *k*-fold cross-validation for the allocation of reference samples between the validation and training datasets is a relatively new concept within remote sensing classification approaches. The ability to determine with certainty that there is no statistically significant change in classification results based on the assignment of training and validation points provides an important degree of certainty in the validity of the classification results. This in turn allows for increased robustness of classification results. Additionally, this process allows for the determination of significant factors affecting classification accuracy. This approach facilitates identification of potential sources of classification error. The applicability of this approach to variables other than the input dataset is broad, and presents many opportunities to identify redundant variables within the classification. This method also allows for optimisation in classification approaches, demonstrated in chapter 2 in which it was shown that classification performance improved for three of the five thematic classes analysed based on the use of a higher spatial resolution WorldView-2 dataset.

- The results of the final chapter present an important case study showing the utility of UAS mounted hyperspectral snap-shot cameras for grassland and other vegetation community differentiation. These sensors have not previously been tested as a potential data source for vegetation community classification, and the success of the analysis undertaken here indicates that they are a highly applicable and useful tool capable of discriminating similar vegetation communities from one another based on spectral properties.

Within the field of object-based image analysis, the contributions to knowledge can be summarised as:

- The creation of a novel method capable of predicting class-specific optimal segmentation scale across multiple input variables. This is an important development with the field of object-based image analysis, and represents the first metric capable of predicting individual optimal segmentation scales for thematic classes. The ability to predict class-specific optimal segmentation scales allows for improved object

delineation, and a reduction in the use of ‘trial-and-error’ approaches to segmentation scale selection. Empirical justification of selected segmentation scales has been a goal within the community for several years, and allows for improvements in classification performance, and a decrease in the subjectivity of segmentation scale selection.

- The ability to assess image segmentations and make predictions of class-specific optimal scale factors using both thematic and spatial attributes of segments is a new advancement in the field of object-based image analysis . The inclusion of thematic components into the assessment of segmentation scale is a novel concept, and has not been applied within an automated approach to optimal segmentation scale prediction before now.

6.6 Limitations and future research

Although the findings of this thesis have made significant contributions to the field of lowland native grassland mapping, there are still limitations to the results obtained. Significant issues surrounding the appropriate scale of reference datasets, relative to sensor spatial resolutions, still need to be resolved. Examples of this included the need to determine whether the level of generalisation present within existing datasets such as the TLC dataset trialled in Chapter 1 is appropriate to act as reference data for expanded community mapping and monitoring. The results indicate that there is now a need for studies incorporating multispectral satellite platforms and spatially detailed field datasets in conjunction. The combination of more detailed field data and multi-spectral approaches will likely yield an improvement from the results seen in this thesis.

The studies undertaken in this thesis are also relatively small scale, especially those in Chapters 3 and 5. The Midlands region of Tasmania is large, and vegetation is highly variable. Additionally, land management practices such as irrigation, grazing and fertilisation are also used across the region, which may result in additional variation within vegetation classes. Land management practices are not accounted for in any chapter of this thesis, and are important considerations within the context of the landscape. Before a broad-scale mapping initiative can be deployed, the effect of these practices on grassland classification accuracy needs to be determined.

The value of multi-temporal datasets for grassland community mapping remains an area that needs to be investigated. Strong seasonal variation has been identified as a key component in grassland community differentiation. This study has not assessed this in any manner, due to the need to first identify the optimal spatial and spectral resolutions required within single-date imagery. The results of this thesis indicate that community differentiation within a single-date data source is possible given appropriate spatial and spectral resolutions, however, the results achieved can still potentially be improved upon significantly, most likely through the use of multi-temporal datasets.

The segmentation algorithm developed in chapter 4 provides a good framework for optimal scale prediction, however, the results of the segmentation assessment cannot inherently

guarantee good classification results. Although segmentation quality is strongly linked to the overall accuracy of classification, final accuracies are also influenced by factors such as the quality of training datasets, and the classification model and parameters employed. Although care was taken to ensure that all input datasets were of high quality, manual digitisation of reference objects is often difficult, particularly in natural environments, due to indistinct class boundaries. The performance of the segmentation assessment method was stronger in the results of case study 1 which was undertaken in an urban environment in which object boundaries are naturally more defined. The results of the various case studies indicate that the approach is indeed applicable in natural landscapes, however, the inherent risk of inaccurately defined reference object needs to be considered when interpreting the results.

6.7 Recommendations

This study has produced several key findings that can be used to improve lowland native grassland mapping within the State of Tasmania. Firstly, this study illustrates that multispectral satellite sensors can provide accurate community maps, and successfully differentiate lowland native grassland communities from both each other and surrounding vegetation types. Additionally, these results were shown to provide significantly higher classification accuracies than the current TASVEG community maps when compared to a coarse-resolution field dataset. The success of these approaches means that community mapping can be semi-automated, and the use of extensive manual digitisation of aerial photography can be reduced. Despite the success of approaches applied within the context of this thesis, there are still several considerations that need to be addressed in future research. Broad-scale multispectral approaches need to be tested much more thoroughly over larger study areas, and should include the use of detailed field datasets obtained at the same spatial resolution as the input datasets. The results of this study indicate the key importance of spatially accurate validation and training datasets, especially when spatial resolution is fine.

The results of the various classification trials employed in this thesis indicate that lowland native grassland communities exhibit distinct spectral characteristics associated with narrow regions of the spectrum, most notably around 550 nm. This indicates that the inclusion of detailed spectral measurements in this region of the spectrum could provide improved differentiation, especially when used in conjunction with multi-temporal analysis. It has also been found that topographic variables are highly important for all vegetation classes, and that there is a need for the inclusion of spatial explicit digital surface models to improve classification results.

6.8 Final Remarks

The findings of this thesis serve as an important first step towards a sustainable mapping scheme for lowland native grassland communities. The studies presented here provide several key findings that can be used as the basis of a method capable of meeting the mandate of increased mapping and monitoring stipulated by the Australian Government as a condition of construction for the Midlands Water Scheme. The mapping approaches employed in this thesis are much faster and easier to repeat than methods relying solely on manual digitisation, and

provide the ability to update maps at a much more frequent interval. Additionally, the ability to predict class-specific segmentation scales can provide advantages within the fields of object-based analysis and vegetation mapping alike, as it provides a means of tailoring the scale of analysis to individual classes.

References

Van Aardt, J., Wynne, R. H., Oderwald, R. G. and Campbell, J. B. (2001) 'Spectral Separability among Six Southern Tree Species', *Photogrammetric Engineering and Remote Sensing*, 67(12), pp. 1–184.

Aasen, H., Bendig, J., Bolten, A., Bennertz, S., Willkomm, G. and Bareth, M. (2014) 'Introduction and preliminary results of a calibration for full-frame hyperspectral cameras to monitor agricultural crops with UAVs', *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-7, p. 5194. doi: 10.5194/isprsarchives-XL-7-1-2014.

Aasen, H., Burkart, A., Bolten, A. and Bareth, G. (2015) 'Generating 3D hyperspectral information with lightweight UAV snapshot cameras for vegetation monitoring : From camera calibration to quality assurance', *ISPRS Journal of Photogrammetry and Remote Sensing*. International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS), 108, pp. 245–259. doi: 10.1016/j.isprsjprs.2015.08.002.

Adjorlolo, C., Mutanga, O., Cho, M. A. and Ismail, R. (2012) 'Spectral resampling based on user-defined inter-band correlation filter: C3 and C4 grass species classification', *International Journal of Applied Earth Observation and Geoinformation*. Elsevier B.V., 21(1), pp. 535–544. doi: 10.1016/j.jag.2012.07.011.

Anderson, K., Gaston, K. J., Anderson, K. and Gaston, K. J. (2013) 'Lightweight unmanned aerial vehicles will revolutionize spatial ecology', *Frontiers in Ecology and the Environment esa*. doi: 10.1890/120150.

Archer, K. and Kimes, R. (2008) 'Empirical characterization of random forest variable importance measures', *Computational Statistics & Data Analysis*, 52(4), pp. 2249–2260.

Baatz, M. and Schäpe, a (2000) 'Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation', *Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symposium Salzburg 2000, Karlsruhe, Herbert Wichmann Verlag*, pp. 12–23. doi: Export Date 6 May 2013.

Becker, B. L., Lusch, D. P. and Qi, J. (2007) 'A classification-based assessment of the optimal spectral and spatial resolutions for Great Lakes coastal wetland imagery', *Remote Sensing of Environment*, 108(1), pp. 111–120. doi: 10.1016/j.rse.2006.11.005.

Beeton, R. (2006) *Advice to the Minister for the Environment, Heritage and the Arts from the Threatened Species Scientific Committee (the Committee) on an Amendment to the List of Threatened Ecological Communities under the Environment Protection and Biodiversity Conservat*. Hobart, Tasmania.

Berni, J. A. J., Zarco-tejada, P. J., Suárez, L. and Fereres, E. (2009) 'Thermal and Narrowband Multispectral Remote Sensing for Vegetation Monitoring From an Unmanned Aerial Vehicle', *IEEE Transactions on Geoscience and Remote Sensing*, 47(3), pp. 722–738.

Blaschke, T. (2010) 'Object based image analysis for remote sensing', *ISPRS Journal of Photogrammetry and Remote Sensing*. Elsevier B.V., 65(1), pp. 2–16. doi: 10.1016/j.isprsjprs.2009.06.004.

Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Queiroz Feitosa, R., van der Meer, F., van der Werff, H., van Coillie, F. and Tiede, D. (2014) 'Geographic Object-Based Image Analysis - Towards a new paradigm', *ISPRS Journal of Photogrammetry and Remote Sensing* 87, pp. 180–191. doi: 10.1016/j.isprsjprs.2013.09.014.

Bock, M., Xofis, P., Mitchley, J., Rossner, G. and Wissen, M. (2005) 'Object-oriented methods for habitat mapping at multiple scales – Case studies from Northern Germany and Wye Downs, UK', *Journal for Nature Conservation*, 13(2–3), pp. 75–89. doi: 10.1016/j.jnc.2004.12.002.

Breckenridge, R. P., Dakins, M., Bunting, S., Harbour, J. L., Lee, R. D., Breckenridge, R. P., Dakins, M., Bunting, S., Harbour, J. L. and Lee, R. D. (2016) 'Using Unmanned Helicopters to Assess Vegetation Cover in Sagebrush Steppe Ecosystems, *Rangeland Ecology and Management*, 65(4), pp. 362–370.

Breiman, L. (2001) 'Random Forests', *Machine Learning*, 45(1), pp. 5–32.

Bruzzone, Lorenzo., Carlin, L. (2006) 'A Multilevel Context-Based System for Classification of Very High Spatial Resolution Images', *IEEE Transactions on Geoscience and Remote Sensing*, 44(9), pp. 2587–2600.

Burai, P., Tomor, T., Beko, L. and Deak, B. (2015) 'Airborne Hyperspectral Remote Sensing For Identification Grassland Vegetation', *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 427–431. doi: 10.5194/isprsarchives-XL-3-W3-427-2015.

Carter, O., Murphy, A. M., Cheal, D., Beecham, B., Bossard, K., Chapman, A., Craigie, V., Denholm, B., Diez, S., Dorrough, J., Dunford, M., Dunn, M., Eddy, D., Fairfax, R., Fensham, R., Ferwerda, F., Foreman, P., Gibson, N. and Gilfedder, L. (2003) *Natural Temperate Grasslands*. Department of Environment and Energy. Canberra, Australia

Chladil, M. A. and Nunez, M. (1995) 'Assessing Grassland Moisture and Biomass in Tasmania- The Application of Remote Sensing and Empirical Models for a Cloudy Environment', *International Journal of Wildland Fire*, 5(3), pp. 165–171.

Cho, M. and Skidmore, K. (2009) 'Hyperspectral predictors for monitoring biomass production in Mediterranean mountain grasslands: Majella National Park, Italy', *International Journal of Remote Sensing*, 30(2), pp. 499–515. doi: 10.1080/01431160802392596.

Cleve, C., Kelly, M., Kearns, F. R. and Moritz, M. (2008) 'Classification of the wildland–urban interface: A comparison of pixel- and object-based classifications using high-resolution aerial photography', *Computers, Environment and Urban Systems*, 32(4), pp. 317–326. doi: 10.1016/j.compenvurbsys.2007.10.001.

Clevers, J. G. P. W., van der Heijden, G. W. a. M., Verzakov, S. and Schaepman, M. E. (2007) 'Estimating grassland biomass using SVM band shaving of hyperspectral data', *Photogrammetric Engineering and Remote Sensing*, 73(9), pp. 1141–1148. doi: 10.14358/PERS.73.10.1141.

Clinton, N., Holt, A., Scarborough, J., Yan, L. and Gong, P. (2010) 'Accuracy assessment measures for object-based image segmentation goodness', *Photogrammetric Engineering & Remote Sensing*, 76(3), pp. 289–299. doi: 10.14358/PERS.76.3.289.

Clinton, N., Holt, A., Yan, L. and Gong, P. (2008) 'An accuracy assessment measure for object based image segmentation', *Internation Archives of Photogrammetry, Remote Sensing and Spatial Information Science*, 4, pp. 1189–1194.

Costa, H., Foody, G. and Boyd, D. (2014) 'Integrating thematic information into the assessment of image segmentation analyses for Object-Based land cover classification', *South-Eastern European Journal of Earth Observation and Geomatics*, 3(25), pp. 155–159.

Cundill, S. L., der van Werff, H. M. A. and der van Meijde, M. (2015) 'Adjusting spectral indices for spectral response function differences of very high spatial resolution sensors simulated from field spectra', *Sensors*, 15(3), pp. 6221–6240. doi: 10.3390/s150306221.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Kyle, T., Gibson, J., Lawler, J. J., Beard, H. and Hess, T. (2007) 'Random Forests for Classification in Ecology', *Ecology*, 88(11), pp. 2783–2792.

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., Corsi, F. and Cho, M. (2008) 'LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements', *ISPRS Journal of Photogrammetry and Remote Sensing*, 63(4), pp. 409–426. doi: 10.1016/j.isprsjprs.2008.01.001.

DEWHA (2010) *Lowland Native Grasslands of Tasmania*. Hobart, Tasmania.

Diaz-Uriarte, R. (2007) 'GeneSrF and varselRF: a web-based tool and R package for gene selection and classification using random forest',*Bioinformatics*, 8(328).

Dilley, A., Millie, S., O'Brien, M. and Edwards, M. (2004) 'The relation between Normalised Difference Vegetatation Index and vegetation moisture content at three grassland locations in Victoria, Australia', *International Journal of Remote Sensing*, 25(19), pp. 3913–3928.

Dorigo, W., Lucieer, A., Podobnikar, T. and Čarni, A. (2012) 'Mapping invasive Fallopia japonica by combined spectral, spatial, and temporal analysis of digital orthophotos', *International Journal of Applied Earth Observation and Geoinformation*, 19, pp. 185–195. doi: 10.1016/j.jag.2012.05.004.

DPIPWE (2010a) *Draft Strategic Assessment for the Water Access Program Midlands Water Scheme , Tasmania Strategic Impact Assessment Report*. Hobart. Tasmania.

DPIPWE (2010b) *Strategic Assessment for the Water Access Program Midlands Water Scheme, Tasmania- Supplementary Report*. Hobart.

DPIPWE (2014) 'TASVEG 3.0-Metadata Statement', pp. 1–22. Hobart, Tasmania.

Dubinin, M., Potapov, P., Lushchekina, A. and Radeloff, V. C. (2010) 'Remote Sensing of Environment Reconstructing long time series of burned areas in arid grasslands of southern Russia by satellite remote sensing', *Remote Sensing of Environment*, 114(8), pp. 1638–1648. doi: 10.1016/j.rse.2010.02.010.

Duro, D. C., Franklin, S. E. and Dubé, M. G. (2012) 'Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests', *International Journal of Remote Sensing*, 33(14), pp. 4502–4526. doi: 10.1080/01431161.2011.649864.

Dusseux, P., Corpetti, T., Hubert-Moy, L. and Corgne, S. (2014) 'Combined Use of Multi-Temoportal Optical and Radar Satellite Images for Grassland Monitoring', *Remote Sensing*, 6, pp. 6163–6182.

Environmental Systems Research Institute (2014) 'ArcGIS Desktop: Release 10.3'. Redlands, California: Environmental Systems Research Insitute.

Exelis Visual Solutions (2013) 'ENVI Version 5.1'. Boulder, Colorado: Harris Geosptatial Solutions.

Fensham, R., Kirkpatrick, J. (1989) 'The conservation of original vegetation remnants in the midlands, tasmania', *Papers and Proceedings from the Royal Society of Tasmania*, 123, pp. 229–246.

Foody, G. M. and Dash, J. (2010) 'Estimating the relative abundance of C 3 and C 4 grasses in the Great Plains from multi-temporal MTCI data: issues of compositing period and spatial generalizability', *International Journal of Remote Sensing*, 31(2), pp. 351–362. doi: 10.1080/01431160902887339.

Gao, J. (2007) 'Quantification of grassland properties: how it can benefit from geoinformatic technologies?', *International Journal of Remote Sensing*, 27(1351–1365).

Gevaert, C. M., Suomalainen, J., Tang, J. and Kooistra, L. (2015) 'Generation of Spectral – Temporal Response Surfaces by Combining Multispectral Satellite and Hyperspectral UAV Imagery for Precision Agriculture Applications', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6), pp. 3140–3146.

Gianelle, D. and Guastella, F. (2016) 'Nadir and off - nadir hyperspectral field data : strengths and limitations in estimating grassland biophysical characteristics', 1161(September). doi: 10.1080/01431160600658180.

Gibbes, C., Adhikari, S., Rostant, L., Southworth, J. and Qiu, Y. (2010) 'Application of Object Based Classification and High Resolution Satellite Imagery for Savanna Ecosystem Analysis', *Remote Sensing*, 2(12), pp. 2748–2772. doi: 10.3390/rs2122748.

Gilfedder, L. (1990) 'Threatened Species From Tasmania's Remnant Grasslands', *Tasforests*, (December 1990), pp. 129–132.

Gitelson, A. and Merzlyak, M. (1994) 'Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer Platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation.', *Journal of Plant Physiology*, 143(3), pp. 286–292.

Goetz, S. (1997) 'Multi-sensor analysis of NDVI, surface temperature and biophysical variables at a mixed grassland site', *International Journal of Remote Sensing*, 18(1).

Graetz, R. D. (1987) 'Satellite Remote Sensing of Australian Rangelands', *Remote Sensing of Environment*. 331, pp. 313–331.

Guerin, G. R., Sparrow, B., Tokmakoff, A., Smyth, A., Leitch, E., Baruch, Z. and Lowe, A. J. (2017) 'Opportunities for Integrated Ecological Analysis across Inland Australia with Standardised Data from Ausplots Rangelands', *PLoS One*. pp. 1–19. doi: 10.4227/05/57C92730DBC06.

Guo, X., Wilmshurst, J., McCanny, S., Fargey, P. and Richard, P. (2004) 'Measuring Spatial and Vertical Heterogeneity of Grasslands Using Remote Sensing Techniques', *Journal of Environmental Informatics*, 3(1).

Haralick, R. M. and Shanmugam, K. (1973) 'Textural Features for Image Classification', *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(6).

Harris, S., Kitchener, A. (2005) 'From Forest to Fjaeldmark: Descriptions of Tasmania's vegetation', Department of Primary Industry, Parks, Water and Environment. Hobart, Tasmania.

Hay, G. J. and Castilla, G. (2006) 'Object-based image analysis: strengths, weaknesses, opportunities and threats (SWOT)', *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, 36, p. 4.

He, Y. and Guo, X. (2006) 'Leaf Area Index estimation using remotely sensed data for Grasslands National Park', *Prairie Perspectives*, 9(1), pp. 105–117.

Irisarri, J. G. N., Oesterheld, M., Verón, S. R. and Paruelo, J. M. (2009) 'Grass species differentiation through canopy hyperspectral reflectance', *International Journal of Remote Sensing*, 30(22), pp. 5959–5975. doi: 10.1080/01431160902791895.

Ji, L. and Peters, A. (2003) 'Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices', *Remote Sensing of Environment*, 87, pp. 85–98.

Jin, Y. X., Xu, B., Yang, X. C., Qin, Z. H., Wu, Q., Zhao, F., Chen, S., Li, J. Y., Ma, H. L., (2015) 'MODIS-based vegetation growth of temperate grassland and its correlation with meteorological factors in northern China', *International Journal of Remote Sensing*, 36. doi: 10.1080/01431161.2015.1079346.

Kawamura, K., Sakuno, Y., Tanaka, Y., Lee, H., Lim, J., Kurokawa, Y. and Watanabe, N. (2011) 'Mapping herbage biomass and nitrogen status in an Italian ryegrass (Lolium multiflorum L.) field using a digital video camera with balloon system', *Journal of Applied Remote Sensing*, 5(1).

Kirkpatrick, J.B., Gilfedder, L. A., Fensham, R. J. (1988) *City Parks and Cemeteries: Tasmania's Remnant Grasslands and Grassy Woodlands*. Hobart, Tasmania: Tasmanian Conservation Trust.

Kitchener, A. and Harris, S. (2013) *From Forest to Fjaeldmark: Descriptions of Tasmania's Vegetation*. Edition 2. Tasmania: Department of Primary Industries, parks, Water and Environment.

Knox, J., Thompson, R. and Campbell, S. (2006) *Species for Profit*. Hobart, Tasmania: Department of Primary Industries, Water and Environment.

Laliberte, A. S., Goforth, M. A., Steele, C. M. and Rango, A. (2011) 'Multispectral Remote Sensing from Unmanned Aircraft : Image Processing Workflows and Applications for Rangeland', *Remote Sensing*, (1), pp. 2529–2551. doi: 10.3390/rs3112529.

Laliberte, A. S., Herrick, J. E., Rango, A. and Winters, C. (2010) 'Acquisition , Or thorectification , and Object-based Classification of Unmanned Aerial Vehicle (UA V) Imager

y for Rangeland Monitoring', *Photogrammetric Engineering & Remote Sensing*, 76(6), pp. 661–672.

Laliberte, A. S., Rango, A. and Herrick, J. (2007) 'Unmanned Aerial Vehicles For Rangeland Mapping And Monitoring : A Comparison Of Two Systems', in *Americal Society of Photogrammetry and Remote Sensing Annual Conference*.

Laws, K. (1980) *Textured Image Segmentation*. University of Southern California.

Liaw, A. and Wiener, M. (2002) 'Classification and Regression by randomForest', *R News*, 2(3), pp. 18–22.

Lu, B., He, Y. and Liu, H. (2016) 'Investigating species composition in a temperate grassland using Unmanned Aerial Vehicle-acquired imagery', in *Fourth International Workshop on Earth Observation and Remote Sensing Applications*.

Lucieer, A. and Stein, A. (2002) 'Existential uncertainty of spatial objects segmented from satellite sensor imagery', *IIEE Transactions on Geoscience and Remote Sensing*, 40(11), pp. 2518–2521. doi: 10.1109/TGRS.2002.805072.

Mansour, K., Mutanga, O., Adam, E. and Abdel-rahman, E. M. (2016) 'Multispectral remote sensing for mapping grassland degradation using the key indicators of grass species and edaphic factors', *Geocarto International*. 31(5), pp. 477–491. doi: 10.1080/10106049.2015.1059898.

Marsett, R., Qui, J., Heilman, P., Biedenbender, S., Watson, M., Amer, S., Weltz, M., Goodrich, D. and Marsett, R. (2006) 'Remote Sensing for Grassland Management in the Arid Southwest', *Rangeland Ecology and Management*, 59(5).

Mcinnes, W. S., Smith, B. and Mcdermid, G. J. (2015) 'Discriminating Native and Nonnative Grasses in the Dry Mixedgrass Prairie With MODIS NDVI Time Series', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(4), pp. 1395–1403.

Mcvicar, T. R. and Jupp, L. B. (1998) 'The Current and Potential Operational Uses of Remote Sensing to Aid Decisions on Drought Exceptional Circumstances in Australia : a Review', *Agricultural Systems*, 57(3), pp. 399–468.

Michaels, K. (2006) *A Manual for Assessing Vegetation Condition in Tasmania*. Version 1. Hobart: Resource Management and Conservation, Department of Primary Industries, Water and Environment.

Möller, M., Lymburner, L. and Volk, M. (2007) 'The comparison index: A tool for assessing the accuracy of image segmentation', *International Journal of Applied Earth Observation and Geoinformation*, 9(3), pp. 311–321. doi: 10.1016/j.jag.2006.10.002.

Mutanga, O. and Skidmore, A. K. (2004) 'Narrow band vegetation indices overcome the saturation problem in biomass estimation', *International Journal of Remote Sensing*, 25(19), pp. 3999–4014. doi: 10.1080/01431160310001654923.

Naidoo, L., Cho, M. a., Mathieu, R. and Asner, G. (2012) 'Classification of savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a Random Forest data mining environment', *ISPRS Journal of Photogrammetry and*

Remote Sensing. International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS), 69, pp. 167–179. doi: 10.1016/j.isprsjprs.2012.03.005.

Neubert, M., Herold, H. and Meinel, G. (2006) 'Evaluation of remote sensing image segmentation quality–further results and concepts', *International Conference on Object-based Image Analysis*, p. 6.

Neubert, M., Herold, H. and Meinel, G. (2006) 'Evaluation Of Remote Sensing Image Segmentation Quality – Further Results And Concepts', *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36.4/C42.

Pal, M. and Foody, G. M. (2010) 'Feature selection for classification of hyperspectral data by SVM', *IEEE Transactions on Geoscience and Remote Sensing*, 48(5), pp. 2297–2307. doi: 10.1109/TGRS.2009.2039484.

Paltridge, G. W. (1988) 'Monitoring Grassland Dryness and Fire Potential in Australia with NOAA / AVHRR Data', *Remote Sensing of Environment.* 394, pp. 381–394.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Wiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011) 'Scikit-Learn: Machine Learning in Python', *Journal of Machine Learning Research*, 12, pp. 2825–2830.

Pickup, G., Basin, G. and Chewings, V. (1994) 'Remote-Sensing-Based Condition Assessment for Nonequilibrium Rangelands Under Large-Scale Commercial Grazing', *Ecological Applications*, 4(3).

Price, K., Guo, X. and Stiles, J. (2002) 'OPtimal Landsat TM band combinations and vegetation indices for distrimination of six grassland types in eastern Kansas', *International Journal of Remote Sensing2*, 23(23).

R Core Team (2015) 'R: A language and environment for statistical computing'. Vienna, Austria: R Foundation for Statistical Computing.

Rahman, A. F., Gamon, J. A., Sims, D. A. and Schmidts, M. (2003) 'Optimum pixel size for hyperspectral studies of ecosystem function in southern California chaparral and grassland', *Remote Sensing of Environment*, 84(2), pp. 192–207. doi: 10.1016/S0034-4257(02)00107-4.

Rango, A., Laliberte, A., Herrick, J., Winters, C., Havstad, K., Steele, C. and Browning, D. (2009) 'Unmanned aerial vehicle-based remote sensing for rangeland assessment, monitoring and management', *Journal of Applied Remote Sensing*, 3(1).

Reed, B. C., Brown, J. F., Vanderzee, D., Loveland, T. R., Merchant, W., Ohlen, D. O., Bradley, C., Jesslyn, F., Thomas, R., James, W. and Donald, O. (2004) 'Measuring Phenological Variability from Satellite Imagery Measuring phenological variability from satellite imagery', *BioScience*, 54(6), pp. 547–560.

Repaka, S.R., Truax, D.D., Kolstad, E.. O'Hara, C. G. (2004) 'Comparing Spectral And Object Based Approaches For Classification And Transportation Feature Extraction From High Resolution Multispectral Imagery', *Proceedings of the American Society of Photogrammetry and Remote Sensing, Denver, May 2004.*

Roth, K. L., Roberts, D. A., Dennison, P. E., Peterson, S. H. and Alonzo, M. (2015) 'The impact of spatial resolution on the classification of plant species and functional types within imaging spectrometer data', *Remote Sensing of Environment*. 171, pp. 45–57. doi: 10.1016/j.rse.2015.10.004.

Rousseeuw, P. J. and Croux, C. (1993) 'Alternatives to the Median Absolute Deviation', *Journal of the American Statistical Association*, 88(424), pp. 1273–1283.

Running, S., Nemani, R., Heinsch, F. A., Zhao, M., Reeves, M. and Hashimoto, H. (2004) 'A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production', *BioScience*, 54(6), pp. 547–560.

Schmidt, K. S. and Skidmore, A. K. (2003) 'Spectral discrimination of vegetation types in a coastal wetland', *Remote Sensing of Environment*, 85(1), pp. 92–108. doi: 10.1016/S0034-4257(02)00196-7.

Seaquist, J., Olsson, L. and Ardo, J. (2003) 'A remote sensing-based primary production model for grassland biomes', *Ecological Modelling*, 169, pp. 131–155.

Silva, B., Lehner, L., Roos, K., Fries, A., Rollenbeck, R., Beck, E. and Bendix, J. (2014) 'Mapping Two Competing Grassland Species from a Low-Altitude Helium Balloon', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(7).

Sim, C. K. and Abdullah, K. (2014) 'Preliminary Research on Grassland Fine-classification Based on MODIS', in *35th International Symposium of Remote Sensing of Environment*. doi: 10.1088/1755-1315/17/1/012079.

Sims, D. A. and Gamon, J. A. (2002) 'Relationships between leaf pigment content and spectral reflectance across a wide range of species , leaf structures and developmental stages', *Remote Sensing of Environment*, 81, pp. 337–354.

Species Section, T. (2006) *Flora Recovery Plan: Tunbridge buttercup, Ranunculus prasinus, 2006-2010*. Hobart, Tasmania.

Svoray, T., Perevolotsky, A. and Atkinson, P. M. (2011) 'Ecological sustainability in rangelands : the contribution of remote sensing', *International Journal of Remote Sensing*, (October17). doi: 10.1080/01431161.2013.793867.

Tieszen, L., Reed, B. C., Bliss, N. B., Wylie, B. K. and DeJong, D. D. (1997) 'NDVI, C3 and C4 Production and Distribution in Great Plains Grassland Land Cover Classes', *Ecological Applications*, 7(1), pp. 59–78.

Tolpekin, V. A. and Stein, A. (2009) 'Quantification of the effects of land-cover-class spectral separability on the accuracy of markov-random-field-based superresolution mapping', *IEEE Transactions on Geoscience and Remote Sensing*, 47(9), pp. 3283–3297. doi: 10.1109/TGRS.2009.2019126.

Trimble Navigation (2014) 'eCognition Developer, Version 9'. Sunnyvale, California: Trimble Navigation.

Tucker, C. J. (1977) 'Use of Near Infrared/Red Radiance Ratios for Estimating Vegetation Biomass and Physiological Status', *NASA/Goddard Space Flight Center*, p. 47.

Tucker, C. J., Justice, C. O. and Prince, S. D. (1986) 'Monitoring the grasslands of the Sahel 1984-1985', *International Journal of Remote Sensing*, 7(11).

Wang, R., Gamon, J. A., Montgomery, R. A., Townsend, P. A., Zygielbaum, A. I., Bitan, K., Tilman, D. and Cavender-bares, J. (2016) 'Seasonal Variation in the NDVI – Species Richness Relationship in a Prairie Grassland Experiment (Cedar Creek)', *Remote Sensing*, 8(2). doi: 10.3390/rs8020128.

Watts, J. D. and Lawrence, R. L. (2008) 'Merging Random Forest Classification with an Object-Oriented Approach for Analysis of Agricultural Lands', *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(B7), pp. 579–582.

Weidner, U. (2008) 'Contribution to the assessment of segmentation quality for remote sensing applications', *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 37(B7), pp. 479–484.

Wen, Q., Zhang, Z., Liu, S., Wang, X. and Wang, C. (2010) 'Classification of Grassland Types by MODIS Time-Series Images in Tibet , China', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(3), pp. 404–409.

Whiteside, T. G., Boggs, G. S. and Maier, S. W. (2011) 'Comparing object-based and pixel-based classifications for mapping savannas', *International Journal of Applied Earth Observation and Geoinformation*, 13(6), pp. 884–893. doi: 10.1016/j.jag.2011.06.008.

Whiteside, T. G., Maier, S. W. and Boggs, G. S. (2014) 'Area-based and location-based validation of classified image objects', *International Journal of Applied Earth Observation and Geoinformation*, 28(1), pp. 117–130.

Williamson, H. (1992) 'Developing a Methodology for Estimating Grassland Variables with Remotely Sensed Data', *The Royal Geographical Society (with the Institute of British Geographers)*, 24(1).

Wood, E. M., Pidgeon, A. M., Radeloff, V. C. and Keuler, N. S. (2012) 'Image texture as a remotely sensed measure of vegetation structure', *Remote Sensing of Environment*. 121, pp. 516–526. doi: 10.1016/j.rse.2012.01.003.

Xu, B., Yang, X. C., Tao, W. G., Miao, J. M., Yang, Z., Liu, H. Q., Jin, Y. X., Zhu, X. H., Qin, Z. H., Lu, H. Y., (2013) 'MODIS-based remote-sensing monitoring of the spatiotemporal patterns of Chinas grassland vegetation growth', *International Journal of Remote Sensing*, 34(11). doi: 10.1080/01431161.2012.762696.

Yan, G., Mas, J., F., Maathuis, B. H. P., Xiangmin, Z. and Van Dijk, P. M. (2006) 'Comparison of pixel based and object oriented image classification approaches—a case study in a coal fire area, Wuda, Inner Mongolia, China', *International Journal of Remote Sensing*, 27(18), pp. 4039–4055. doi: 10.1080/01431160600702632.

Yang, J., Li, P. and He, Y. (2014) 'A multi-band approach to unsupervised scale parameter selection for multi-scale image segmentation', *ISPRS Journal of Photogrammetry and Remote Sensing*. International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS), 94, pp. 13–24. doi: 10.1016/j.isprsjprs.2014.04.008.

Yang, X., Guo, X. and Fitzsimmons, M. (2012) 'Assessing light to moderate grazing effects

on grassland production using satellite imagery', *International Journal of Remote Sensing*, 33(16).

Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M. and Schirokauer, D. (2006) 'Object-based Detailed Vegetation Classification with Airborne High Spatial Resolution Remote Sensing Imagery', *Photogrammetric Engineering and Remote Sensing*, 72(7), pp. 799–811.

Zacharek, A., Gilfedder, L. and Harris, S. (1997) 'The flora of Township Lagoon Nature Reserve and its management, Tunbridge, Tasmania', *Papers and Proceedings of the Royal Society of Tasmania*, 131(Davies 1988), pp. 57–66.

Zarco-Tejada, P. J., Gonzalez-Dugo, V. and Berni, J. A. J. (2012) 'Remote Sensing of Environment Fluorescence , temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera', *Remote Sensing of Environment*. 117, pp. 322–337. doi: 10.1016/j.rse.2011.10.007.

Zarco-tejada, P. J., Guillén-climent, M. L., Hernández-clemente, R. and Catalina, A. (2013) 'Agricultural and Forest Meteorology Estimating leaf carotenoid content in vineyards using high resolution hyperspectral imagery acquired from an unmanned aerial vehicle (UAV)', *Agricultural and Forest Meteorology*.171–172, pp. 281–294. doi: 10.1016/j.agrformet.2012.12.013.

Zarco-Tejada, P. j, Ustin, S. L. and ML, W. (2005) 'Temporal and Spatial Relationships between Within-Field Yield Variability in Cotton and High Spatial Hyperspectral Remote Sensing Imagery', *Agronomy Journal*, 97(3), pp. 641–653. doi: 10.2134/agronj2003.0257.

Zerger, A., McIntyre, S., Gobbett, D. and Stol, J. (2011) 'Remote detection of grassland nutrient status for assessing ground layer vegetation condition and restoration potential of eucalypt grassy woodlands', *Landscape and Urban Planning*.

Zhan, Q., Molenaar, M., Tempfli, K. and Shi, W. (2005) 'Quality assessment for geo-spatial objects derived from remotely sensed data', *International Journal of Remote Sensing*, 26(14), pp. 2953–2974. doi: 10.1080/01431160500057764.

Zhang, X., Friedl, M. a., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., Reed, B. C. and Huete, A. (2003) 'Monitoring vegetation phenology using MODIS', *Remote Sensing of Environment*, 84(3), pp. 471–475. doi: 10.1016/S0034-4257(02)00135-9.

Zur, Y., Gitelson, A., Chivkunova, O. and Merzlyak, M. (2000) 'The spectral contribution of carotenoids to light absorption and reflectance in green leaves', *Papers in Natural Resources*, 2(272).